

154679

STIC-Biotech/ChemLib

From: Whiteman, Brian  
Sent: Friday, May 27, 2005 9:45 AM  
To: STIC-Biotech/ChemLib  
Subject: seq search

09/610,313  
Barnett et al.

SEQ ID NOs: 30, 31, and 32  
1) nucleotide search against databases  
2) amino acid search against databases

Thank you,

Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

RECEIVED  
MAY 27 2005  
(STIC)

5

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2-\_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: 6/3 \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA#: \_\_\_\_\_ AA#: \_\_\_\_\_  
Interference: \_\_\_\_\_ SPDI: \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure#: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2005, 23:20:00 ; Search time 1270.42 Seconds  
(without alignments)  
11504.726 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469

Sequence: 1 gtcagccaccatggccga.....gggtagcaccgtgaattc 2469

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2469	100.0	2469	6	ABL39959 Synthetic
2	2469	100.0	2469	12	Adm73764 HIV-1 pol
3	2457	99.5	2457	8	ACA03548 Synthetic
4	2457	99.5	2457	10	Adc13266 DNA of HI
5	2442.2	98.9	2463	6	ABL39960 Synthetic
6	2442.2	98.9	2463	12	Adm73765 HIV-1 pol
7	2434.6	98.6	2457	8	ACA03547 Synthetic
8	2434.6	98.6	2457	10	Adc13265 DNA of HI
9	2415.4	97.8	2457	6	ABL39961 Synthetic
10	2415.4	97.8	2457	12	Adm73766 HIV-1 pol
11	2401.8	97.3	2445	8	ACA03546 Synthetic
12	2401.8	97.3	2445	10	Adc13264 DNA of HI
13	2394.8	97.0	3930	10	Adc13230 DNA of HI
14	2393.2	96.9	3930	10	Adc13231 DNA of HI
15	2393.2	96.9	3930	10	Adc13232 DNA of HI
16	2393.2	96.9	5184	8	ACA03591 Synthetic
17	2393.2	96.9	5184	10	Adc13279 DNA of HI
18	2362.8	95.7	3531	10	Adc13234 DNA of HI
19	2361.2	95.6	3537	10	Adc13236 DNA of HI
20	2360.2	95.6	5145	8	ACA03521 Synthetic

## ALIGNMENTS

### RESULT 1

ABL39959

ID ABL39959 standard; DNA; 2469 BP.

AC ABL39959;

XX 15-MAY-2002 (first entry)

DT 15-MAY-2002 (first entry)

XX Synthetic construct PR975(+) SEQ ID NO:30.

DE Human immunodeficiency virus type C; antigenic HIV type C protein; immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef; immunostimulant; Gene therapy; gene; ds.

XX Human immunodeficiency virus; type C.

OS Synthetic.

XX WO200204493-A2.

PN 17-JAN-2002.

XX 05-JUL-2001; 2001WO-US021241.

PF 05-JUL-2000; 2000US-00610313.

PR (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2002-154920/20.

DR New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.

PT Claim 1; Fig 8; 233pp; English.

XX The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef (i). (i) have immunostimulant activity and can be used in gene therapy.

21 2360.2 95.6 5145 10 ADC13233  
22 2350.2 95.2 3538 10 ADC13235  
23 2349.4 95.2 3624 8 ACA03550  
24 2349.4 95.2 3624 10 ADC13268  
25 2301.8 93.2 3607 8 ACA03551  
26 2301.8 93.2 3607 10 ADC13269  
27 2283.6 92.5 3597 8 ACA03549  
28 2283.6 92.5 3597 10 ADC13267  
29 2149.6 87.1 2472 8 ACA03543  
30 2149.6 87.1 2472 8 ACC78507  
31 2121.2 85.9 2466 8 ACA03542  
32 2121.2 85.9 2466 8 ACC78506  
33 2094.4 84.8 2460 8 ACA03541  
34 2094.4 84.8 2460 8 ACC78505  
35 2093.4 84.8 3564 8 ACC78488  
36 2093.4 84.8 3564 8 ACC78489  
37 2092.8 84.8 4716 8 ACA03522  
38 2092.8 84.8 4716 10 ADC13238  
39 2089 84.6 3999 8 ACC78484  
40 2087.4 84.5 3999 8 ACC78485  
41 2087.4 84.5 3999 8 ACC78486  
42 2087.4 84.5 5283 8 ACA03584  
43 2087.4 84.5 5283 8 ACC78529  
44 2087.2 84.5 4713 8 ACA03592  
45 2087.2 84.5 4713 10 ADC13280

ADC13233 DNA of HI  
ADC13235 DNA of HI  
ACA03550 Synthetic  
ADC13268 DNA of HI  
ACA03551 Synthetic  
ADC13269 DNA of HI  
ACA03549 Synthetic  
ADC13267 DNA of HI  
ACC78507 HIV p2Pol  
ACA03542 Synthetic  
ACC78506 HIV p2Pol  
ACA03541 Synthetic  
ACC78505 HIV p2Pol  
ACC78488 HIV GagPo  
ACC78489 HIV GagPo  
ACA03522 Synthetic  
ADC13238 DNA of HI  
ACC78484 HIV GagCo  
ACC78485 HIV GagCo  
ACC78486 HIV GagCo  
ACA03584 Synthetic  
ACC78529 HIV TatRe  
ACA03592 Synthetic  
ADC13280 DNA of HI

CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC Type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 100.0%; Score 2469; DB 6; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 7.4e-298;  
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAGCCACCATGCGCGAGGCATGAGCAGGCCACACAGCCAGCCCAACATCCTGATGTCAG 60  
DB 1 GTCGAGCCACCATGCGCGAGGCATGAGCAGGCCACACAGCCACCAACATCCTGATGTCAG 60

QY 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 120  
DB 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 120

QY 121 CACATGCGCGCAACTGCGCGAGGCATGAGCAGGCCACACAGCCACCAACATCCTGATGTCAG 180  
DB 121 CACATGCGCGCAACTGCGCGAGGCATGAGCAGGCCACACAGCCACCAACATCCTGATGTCAG 180

QY 181 GGCACACAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240  
DB 181 GGCACACAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240

QY 241 TTCCCGCAGGCAAGGCCCGCGAGTTTCCAGCGAGCAGAACCCGCGCAACAGCCCCACC 300  
DB 241 TTCCCGCAGGCAAGGCCCGCGAGTTTCCCGAGCAGAACCCGCGCAACAGCCCCACC 300

QY 301 AGCGCGAGTGCAGGTGCGCGGAGCAACCCCGCGAGCGAGCCCGCGCGCGAGCGCCAG 360  
DB 301 AGCGCGAGTGCAGGTGCGCGGAGCAACCCCGCGAGCGAGCCCGCGCGCGAGCGCCAG 360

QY 361 GGCACCTGAACCTTCCCGAGATCACCTGTGCGAGCGCCCTTGTGAGCATCAAGGTG 420  
DB 361 GGCACCTGAACCTTCCCGAGATCACCTGTGCGAGCGCCCTTGTGAGCATCAAGGTG 420

QY 421 GCGCGCCAGATCAAGGAGCCCTGCTGACACCGCGCGCGAGCACCTGTGCGAGGAG 480  
DB 421 GCGCGCCAGATCAAGGAGCCCTGCTGACACCGCGCGCGAGCACCTGTGCGAGGAG 480

QY 481 ATGAGCTGCGCGAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG 540  
DB 481 ATGAGCTGCGCGAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG 540

QY 541 GTGCGCAGTACGACCATGATCTGATCGAGATCTGCGGCAAGAGGCCATCGCGCCCGTG 600  
DB 541 GTGCGCAGTACGACCATGATCTGATCGAGATCTGCGGCAAGAGGCCATCGCGCCCGTG 600

QY 601 CTGATCGGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC 660  
DB 601 CTGATCGGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC 660

QY 661 ACCCTGAATCTTCCCATCAGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGCATG 720  
DB 661 ACCCTGAATCTTCCCATCAGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGCATG 720

QY 721 GAGCGCCCAAGGTGAAGCAGTGGCCCTCACCGAGGAGAGATCAAGGCCCTTGAACCGCC 780  
DB 721 GAGCGCCCAAGGTGAAGCAGTGGCCCTCACCGAGGAGAGATCAAGGCCCTTGAACCGCC 780

QY 781 ATCTGCGAGGAGTGAAGAGGAGGCAAGATCAACAAGATCGCGCCCGCGAGAACCCCTAC 840  
DB 781 ATCTGCGAGGAGTGAAGAGGAGGCAAGATCAACAAGATCGCGCCCGCGAGAACCCCTAC 840

QY 841 AACACCCCGTGTTCGCGCATCAAGAGAGAGCAGCAGCAGTGGCGCGAGCTGGTGGAC 900  
DB 841 AACACCCCGTGTTCGCGCATCAAGAGAGAGCAGCAGCAGTGGCGCGAGCTGGTGGAC 900

QY 901 TTCCGCGAGTGTAACAAAGCGCACCCAGGACTTCTGGAGAGTGAGCTGGGCAATCCCCAC 960  
DB 901 TTCCGCGAGTGTAACAAAGCGCACCCAGGACTTCTGGAGAGTGAGCTGGGCAATCCCCAC 960

QY 961 CCGCGCGGCTGAAAGAAAGAGAGCGTGACCGTGTGAGACGTGGCGAGCGCTACTTC 1020  
DB 961 CCGCGCGGCTGAAAGAAAGAGAGCGTGACCGTGTGAGACGTGGCGAGCGCTACTTC 1020

QY 1021 AGCGTCCCTGGACGAGGACTTCCGCAAGTACACCGCTTACCGCTTCCAGCATCCCC 1080  
DB 1021 AGCGTCCCTGGACGAGGACTTCCGCAAGTACACCGCTTACCGCTTCCAGCATCCCC 1080

QY 1081 AACGAGACCCCGGCAATCCGCTACCAAGTACAAAGTGTGCGCCAGGCTGGAAGGGCAGC 1140  
DB 1081 AACGAGACCCCGGCAATCCGCTACCAAGTACAAAGTGTGCGCCAGGCTGGAAGGGCAGC 1140

QY 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGGCCGCAACCCC 1200  
DB 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGGCCGCAACCCC 1200

QY 1201 GAGATCGTGTATCTACCAAGTACATGAGCAGCTGTAGTGGGAGCGACCTGGAGATCGGC 1260  
DB 1201 GAGATCGTGTATCTACCAAGTACATGAGCAGCTGTAGTGGGAGCGACCTGGAGATCGGC 1260

QY 1261 CAGCACCGGCCCAAGATCGAGGAGCTGCGCAAGCAGCTGTGCGCTGGGGCTTCCACACC 1320  
DB 1261 CAGCACCGGCCCAAGATCGAGGAGCTGCGCAAGCAGCTGTGCGCTGGGGCTTCCACACC 1320

QY 1321 CCGCAAGAGACACAGAAAGAGGCCCTTCTGTGTGATGGGTACGAGCTGCACCCC 1380  
DB 1321 CCGCAAGAGACACAGAAAGAGGCCCTTCTGTGTGATGGGTACGAGCTGCACCCC 1380

QY 1381 GACAAAGTGACCGTGCAGCCCATCGAGCTGCGCGAGAGGAGAGCTTGAACGAGCAGC 1440  
DB 1381 GACAAAGTGACCGTGCAGCCCATCGAGCTGCGCGAGAGGAGAGCTTGAACGAGCAGC 1440

QY 1441 ATCCAGAAGCTGTGGGCAAGCTGAATCTGGGCGAGCAGATCTACCCCGCATCAAGGTG 1500  
DB 1441 ATCCAGAAGCTGTGGGCAAGCTGAATCTGGGCGAGCAGATCTACCCCGCATCAAGGTG 1500

QY 1501 CGCAGCTGTGAAGCTGTGCGCGGCGCAAGGCCCTGACCGACATCTGTCGCTGACCC 1560  
DB 1501 CGCAGCTGTGAAGCTGTGCGCGGCGCAAGGCCCTGACCGACATCTGTCGCTGACCC 1560

QY 1561 GAGGAGCGCGAGCTGAGCTGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGCG 1620  
DB 1561 GAGGAGCGCGAGCTGAGCTGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGCG 1620

QY 1621 GTGTACTACGACCCCAAGGACCTGGTGGCCGAGATCCAGAAAGCAGGCGCCACGACAG 1680  
DB 1621 GTGTACTACGACCCCAAGGACCTGGTGGCCGAGATCCAGAAAGCAGGCGCCACGACAG 1680

QY 1681 TGAACCTACAGATCTACGAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCAAG 1740  
DB 1681 TGAACCTACAGATCTACGAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCAAG 1740

QY 1741 ATGCGCACCGCCCAACACCAAGCAGCTGACCGAGGCGCGTGCAGAAAGATCGCC 1800  
DB 1741 ATGCGCACCGCCCAACACCAAGCAGCTGACCGAGGCGCGTGCAGAAAGATCGCC 1800

QY 1801 ATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAGGAG 1860  
DB 1801 ATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAGGAG 1860

QY 1861 ACCTGGAGACCTGGTGGACCGGACTACTGGCAGGCCACCTGGATCCCCGAGTGGAGTTC 1920  
DB 1861 ACCTGGAGACCTGGTGGACCGGACTACTGGCAGGCCACCTGGATCCCCGAGTGGAGTTC 1920

QY 1921 GTCAACACCCCGCTTGTGTAAGCTGTGGTACACGCTGGAGAAAGAGCCCATCATCGGC 1980  
DB 1921 GTCAACACCCCGCTTGTGTAAGCTGTGGTACACGCTGGAGAAAGAGCCCATCATCGGC 1980

QY 1981 GCCGAGACCTTCTACGTTGAGACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2040



```
QY 781 ATCTGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGSCCCCGAGAACCCCTAC 840
DB |||||
QY 781 ATCTGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGSCCCCGAGAACCCCTAC 840
DB |||||
QY 841 AACACCCCGCTGTTCGCCATCAAGAAGAAGACAGCAACCAAGTGGCGCAAGCTGGTGGAC 900
DB |||||
QY 841 AACACCCCGCTGTTCGCCATCAAGAAGAAGACAGCAACCAAGTGGCGCAAGCTGGTGGAC 900
DB |||||
QY 901 TTCGGGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCGCCAC 960
DB |||||
QY 901 TTCGGGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCGCCAC 960
DB |||||
QY 961 CCGCGCGCTGGAAGAAGAAGAGAGCTGACCGTGTGACAGTGGCGCAGCGCTACTTC 1020
DB |||||
QY 961 CCGCGCGCTGGAAGAAGAAGAGAGCTGACCGTGTGACAGTGGCGCAGCGCTACTTC 1020
DB |||||
QY 1021 AGCGTGCCCTGGACGAGGACTTTCGCAAGTACACCGCTTCAACATCCCCAGCATCAAC 1080
DB |||||
QY 1021 AGCGTGCCCTGGACGAGGACTTTCGCAAGTACACCGCTTCAACATCCCCAGCATCAAC 1080
DB |||||
QY 1081 AACGAGACCCCGGCATCCGCTACCAAGTACACCGCTTCAACATCCCCAGCATCAAC 1140
DB |||||
QY 1081 AACGAGACCCCGGCATCCGCTACCAAGTACACCGCTTCAACATCCCCAGCATCAAC 1140
DB |||||
QY 1141 CCAGCATCTTCAGAGCAGATGACCAAGATCTGGAGCCCTTCGCGCCCGCAACCCC 1200
DB |||||
QY 1141 CCAGCATCTTCAGAGCAGATGACCAAGATCTGGAGCCCTTCGCGCCCGCAACCCC 1200
DB |||||
QY 1201 GAGATCGTGATCTACCAAGTACATGAGCAGCTGTGACGCGGACGACCTTGGAGATCGGC 1260
DB |||||
QY 1201 GAGATCGTGATCTACCAAGTACATGAGCAGCTGTGACGCGGACGACCTTGGAGATCGGC 1260
DB |||||
QY 1261 CAGCACCGCGCAGATTCGAGAGCTCGCGAAGCACTGTGCGCTGGGCTTCAACACC 1320
DB |||||
QY 1261 CAGCACCGCGCAGATTCGAGAGCTCGCGAAGCACTGTGCGCTGGGCTTCAACACC 1320
DB |||||
QY 1321 CCGACAAGAAGACCAAGAGGAGCCCTTCTGTGGATGGCTACGAGCTGCAACCC 1380
DB |||||
QY 1321 CCGACAAGAAGACCAAGAGGAGCCCTTCTGTGGATGGCTACGAGCTGCAACCC 1380
DB |||||
QY 1381 GACAAGTGAACCGGTGACGCCATCGAGCTGCCGAGAGAGCTGAGCCGTGAACGAC 1440
DB |||||
QY 1381 GACAAGTGAACCGGTGACGCCATCGAGCTGCCGAGAGAGCTGAGCCGTGAACGAC 1440
DB |||||
QY 1441 ATCCAGAGCTGTGGCAAGCTGACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG 1500
DB |||||
QY 1441 ATCCAGAGCTGTGGCAAGCTGACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG 1500
DB |||||
QY 1501 CGCAGCTGTGCAAGCTGTGCGCGGCCAAGGCCCTGACGACATCGTCCCTGACC 1560
DB |||||
QY 1501 CGCAGCTGTGCAAGCTGTGCGCGGCCAAGGCCCTGACGACATCGTCCCTGACC 1560
DB |||||
QY 1561 GAGGAGCCGAGCTGGAGCTGGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGGC 1620
DB |||||
QY 1561 GAGGAGCCGAGCTGGAGCTGGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGGC 1620
DB |||||
QY 1621 GTGTACTACGACCCGACAGAGACTGTGGTGGCGAGATCCAGAAGCAAGGCGCAGACAG 1680
DB |||||
QY 1621 GTGTACTACGACCCGACAGAGACTGTGGTGGCGAGATCCAGAAGCAAGGCGCAGACAG 1680
DB |||||
QY 1681 TGGACCTTACCAGATCTTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCCAAG 1740
DB |||||
QY 1681 TGGACCTTACCAGATCTTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCCAAG 1740
DB |||||
QY 1741 ATGCGCACCCGCCACACCAAGCAGTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCC 1800
DB |||||
QY 1741 ATGCGCACCCGCCACACCAAGCAGTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCC 1800
DB |||||
QY 1801 ATGAGAGCATGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1860
DB |||||
QY 1801 ATGAGAGCATGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1860
DB |||||
```

```
QY 1861 ACCTGGGAGACCTGTGGAGCCGACTACTGCGAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
DB |||||
QY 1861 ACCTGGGAGACCTGTGGAGCCGACTACTGCGAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
DB |||||
QY 1921 GTGAACACCCCGCTGTGTGTAAGCTGTGGTACAGTGGAGAACGAGCCCATCATCGGC 1980
DB |||||
QY 1921 GTGAACACCCCGCTGTGTGTAAGCTGTGGTACAGTGGAGAACGAGCCCATCATCGGC 1980
DB |||||
QY 1981 GCCGAGACCTTCTACCTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2040
DB |||||
QY 1981 GCCGAGACCTTCTACCTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2040
DB |||||
QY 2041 TACGTGACCCGACCGCGCGCCCGCAGAGATCGTGAGCTTGAAGAGACCAACCAAG 2100
DB |||||
QY 2041 TACGTGACCCGACCGCGCGCGCCCGCAGAGATCGTGAGCTTGAAGAGACCAACCAAG 2100
DB |||||
QY 2101 ACCGAGTGCAGGCGCATTCAGCTGGCCCTGCGAGGAAGCGGCGAGCGAGGTGAACATCGTG 2160
DB |||||
QY 2101 ACCGAGTGCAGGCGCATTCAGCTGGCCCTGCGAGGAAGCGGCGAGCGAGGTGAACATCGTG 2160
DB |||||
QY 2161 ACCGAGCGCAGTACCGCTTGGCCATTCAGGCGCGAGCCCGACCAAGAGCGAGAGCGAG 2220
DB |||||
QY 2161 ACCGAGCGCAGTACCGCTTGGCCATTCAGGCGCGAGCCCGACCAAGAGCGAGAGCGAG 2220
DB |||||
QY 2221 CTGGTGAACCAAGGATTCAGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2280
DB |||||
QY 2221 CTGGTGAACCAAGGATTCAGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2280
DB |||||
QY 2281 CCCGCCCAACAAAGGATTCGCGGCAACGAGCAGATCGAACAGCTGGTGAAGGGGATC 2340
DB |||||
QY 2281 CCCGCCCAACAAAGGATTCGCGGCAACGAGCAGATCGAACAGCTGGTGAAGGGGATC 2340
DB |||||
QY 2341 CGCAAGTGTCTTCTCGGACCGCTGATGGCGCATCTGATCTACCAAGTACATCGAC 2400
DB |||||
QY 2341 CGCAAGTGTCTTCTCGGACCGCTGATGGCGCATCTGATCTACCAAGTACATCGAC 2400
DB |||||
QY 2401 GACTGTACGTGGCGAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2460
DB |||||
QY 2401 GACTGTACGTGGCGAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2460
DB |||||
QY 2461 GGTGAATTC 2469
DB |||||
QY 2461 GGTGAATTC 2469
DB |||||

RESULT 3
ACA03548
ID ACA03548 standard; DNA; 2457 BP.
XX
AC ACA03548;
XX
AC ACA03548;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA encoding immunogenic HIV peptide #31.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-031680P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
```



```
Db 1681 TACCAGATCTACCAAGGAGCCCTTCAAGAACTGAGAACCGGCAAGTACGCGCAAGATCGCG 1740
Qy 1747 ACCGCCACACCAACGACGTGAAGACGTACCGAGGCGCGTGCAGAGATCGCCATGGAG 1806
Db 1741 ACCGCCACACCAACGACGTGAAGACGTACCGAGGCGCGTGCAGAGATCGCCATGGAG 1800
Qy 1807 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAGAGGACCTGG 1866
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAGAGGACCTGG 1860
Qy 1867 GAGACTGTGTGGACCGCACTACTGCGAGGCCACCTGGATCCCGAGTGGGATTCGTGAAC 1926
Db 1861 GAGACTGTGTGGACCGCACTACTGCGAGGCCACCTGGATCCCGAGTGGGATTCGTGAAC 1920
Qy 1927 ACCCCCCCTGTGTGAAGCTGTGTGTACCACTGGAGAGAGGCCCATCATCGGCGCGGAG 1986
Db 1921 ACCCCCCCTGTGTGAAGCTGTGTGTACCACTGGAGAGAGGCCCATCATCGGCGCGGAG 1980
Qy 1987 ACCTTCTACGTGGACGCGCGCCCAACCGCGAGACCAAGATCGCAAGGCGCGTACGTG 2046
Db 1981 ACCTTCTACGTGGACGCGCGCCCAACCGCGAGACCAAGATCGCAAGGCGCGTACGTG 2040
Qy 2047 ACCGACCGGGCGCGCAAGATGTGTAGCTGACCGCTGACCGAGACCAACCAAGAGACCGAG 2106
Db 2041 ACCGACCGGGCGCGCAAGATGTGTAGCTGACCGCTGACCGAGACCAACCAAGAGACCGAG 2100
Qy 2107 CTGAGGCCATTCAGCTGGCGCTGCAGGACGCGGCGAGGAGTGAACATCGTACCGGAC 2166
Db 2101 CTGAGGCCATTCAGCTGGCGCTGCAGGACGCGGCGAGGAGTGAACATCGTGAACCGAC 2160
Qy 2167 AGCCAGTACGCTGGGCGCATCTCCAGGCGCGAGCCGACCAAGAGCGAGCGAGCTGGTG 2226
Db 2161 AGCCAGTACGCTGGGCGCATCTCCAGGCGCGAGCCGACCAAGAGCGAGCGAGCTGGTG 2220
Qy 2227 AACGAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGCCCGCC 2286
Db 2221 AACGAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGCCCGCC 2280
Qy 2287 CACAGGGGATCGGCGGCAAGCAGCAGATCGAAGCTGTGTGACGAGCGGCGATCCGCGAG 2346
Db 2281 CACAGGGGATCGGCGGCAAGCAGCAGATCGAAGCTGTGTGACGAGCGGCGATCCGCGAG 2340
Qy 2347 GTGCTGTTCCTGGAACGCGATCGATGGCGGCGATCGTGATCTACAGTACATGACGACCTG 2406
Db 2341 GTGCTGTTCCTGGAACGCGATCGATGGCGGCGATCGTGATCTACAGTACATGACGACCTG 2400
Qy 2407 TAGTGGGACAGCGCGCCCTTAGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 2463
Db 2401 TAGTGGGACAGCGCGCCCTTAGATCGATTAAAGCTTCCGGGGCTAGCACCGGT 2457

RESULT 4
ADCI3266
ID ADCI3266 standard; DNA; 2457 BP.
AC
AC ADCI3266;
XX
XX 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
XX Human immunodeficiency virus.
XX
XX WO2003004620-A2.
XX
XX 16-JAN-2003.
XX
XX 05-JUL-2002; 2002WO-US021420.
XX
XX 05-JUL-2001; 2001US-0303192P.
XX
```

```
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX (CHIR ) CHIRON CORP.
XX (UYST-) UNIV STELLENBOSCH.
XX
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX WPI; 2003-221593/21.
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX Prot, or Rev polypeptide, useful for immunization, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 42; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intramusosally, intranasally,
XX subcutaneously, intradermally, transdermally, intravaginally,
XX intrarectally, orally or intravenously. The expression cassette is useful
XX for immunisation, generating packaging cell lines and producing HIV
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX type C related sequence of the invention.
XX
XX Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;
XX
XX Query Match 99.5%; Score 2457; DB 10; Length 2457;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-296;
XX Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 7 GCCACCATGCGCGAGGCCATGAGCCAGGCCAACCGAGGCCAACATCTCTGATGCGAGCGCAGC 66
Db 1 GCCACCATGCGCGAGGCCATGAGCCAGGCCAACCGAGGCCAACATCTCTGATGCGAGCGCAGC 60
Qy 67 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCACTGCGCGAGGAGGGCCACATC 126
Db 61 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCACTGCGCGAGGAGGGCCACATC 120
Qy 127 GCCCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186
Db 121 GCCCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Qy 187 CAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCC 246
Db 181 CAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCC 240
Qy 247 CAGGGCAAGGCCCGCGAGTTCCTCCAGGAGCAGAACCGCGCCCAAGCCCCCAGCGCGC 306
Db 241 CAGGGCAAGGCCCGCGAGTTCCTCCAGGAGCAGAACCGCGCCCAAGCCCCCAGCGCGC 300
Qy 307 GAGCTCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db 301 GAGCTCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Qy 367 CTGAACCTTCCCCCAGATCACTCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
Db 361 CTGAACCTTCCCCCAGATCACTCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Qy 427 CAGATCAAGGAGGCCCTTCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486
Db 421 CAGATCAAGGAGGCCCTTCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Qy 487 CTGCCCGCAAGTGGAAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546
Db 481 CTGCCCGCAAGTGGAAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Qy 547 CAGTACGACCGAGATCTCTGATTCGAGATCTGCGGCAAGAGGCCATCGGACCGTCTGCTGATC 606
Db 547 CAGTACGACCGAGATCTCTGATTCGAGATCTGCGGCAAGAGGCCATCGGACCGTCTGCTGATC 606
```



```
Db 541 CAGTACGACAGATCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATC 600
Qy 607 GGGCCCAACCCCGTGAAACATCATCGGCCGCAACATGTGACCCCAAGCTGGGCTGACCCCTG 666
Db 601 GGGCCCAACCCCGTGAAACATCATCGGCCGCAACATGTGACCCCAAGCTGGGCTGACCCCTG 660
Qy 667 AACTTCCCATCAGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGCATGACCGGC 726
Db 661 AACTTCCCATCAGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGCATGACCGGC 720
Qy 727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGC 786
Db 721 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGC 780
Qy 787 GAGGAGATGGAGAGAGGGGCAAGATCAACAGATTCGGCCCGGAGAACCCCTCAACAAC 846
Db 781 GAGGAGATGGAGAGAGGGGCAAGATCAACAGATTCGGCCCGGAGAACCCCTCAACAAC 840
Qy 847 CCCGTGTTCCGCATCAAGAGAGAGGACAGCAGCAGTGGCGCAAGCTGGTGGATTCGC 906
Db 841 CCCGTGTTCCGCATCAAGAGAGAGGACAGCAGCAGTGGCGCAAGCTGGTGGATTCGC 900
Qy 907 GAGCTGAACAGCGCACCCAGAGCTTCTGGGAGGTGCAGCTGGGCATCCGCCACCCCGCC 966
Db 901 GAGCTGAACAGCGCACCCAGAGCTTCTGGGAGGTGCAGCTGGGCATCCGCCACCCCGCC 960
Qy 967 GGCTGTAAGAGAGAGAGAGCGTGACCGTGTGAGCGTGGCGCAGCGCTACTTCAGGGTG 1026
Db 961 GGCTGTAAGAGAGAGAGAGCGTGACCGTGTGAGCGTGGCGCAGCGCTACTTCAGCGTG 1020
Qy 1027 CCCCTGAGCAGGAGCTTCGCGAAGTACACCGCTTACCATCCCGAGGATCAACAACGAG 1086
Db 1021 CCCCTGAGCAGGAGCTTCGCGAAGTACACCGCTTACCATCCCGAGGATCAACAACGAG 1080
Qy 1087 ACCCCCGGCATCCGCTACAGTACAGCTGCTGCCCGCAGGGCTGAGAGGGCAGCCCGAGC 1146
Db 1081 ACCCCCGGCATCCGCTACAGTACAGCTGCTGCCCGCAGGGCTGAGAGGGCAGCCCGAGC 1140
Qy 1147 ATCTTCCAGAGCAGCAGTACCAAGATCCTGAGAGCCCTTCGCGCGCCGCAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGCAGTACCAAGATCCTGAGAGCCCTTCGCGCGCCGCAACCCCGAGATC 1200
Qy 1207 GTGATCTACAGTACATGAGCAGCTGTGTCGGGAGCGACCTGAGAGATCGGCGCAGGAC 1266
Db 1201 GTGATCTACAGTACATGAGCAGCTGTGTCGGGAGCGACCTGAGAGATCGGCGCAGGAC 1260
Qy 1267 CGGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTACACACCCCGGAC 1326
Db 1261 CGGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTACACACCCCGGAC 1320
Qy 1327 AAGAAGCACCAAGAGGAGCCCGCTTCTGTGGATGGCTACGAGCTGCACCCCGCAAG 1386
Db 1321 AAGAAGCACCAAGAGGAGCCCGCTTCTGTGGATGGCTACGAGCTGCACCCCGCAAG 1380
Qy 1387 TGGACCTGACGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1446
Db 1381 TGGACCTGACGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy 1447 AAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGCGCCAG 1506
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy 1507 CTGTGAAGCTGTGCGCGGCGCAAGCCCTGACCGACATCGTGGCCCTGACCGAGGAG 1566
Db 1501 CTGTGAAGCTGTGCGCGGCGCAAGCCCTGACCGACATCGTGGCCCTGACCGAGGAG 1560
Qy 1567 GCGGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACCGCGTGTAC 1626
Db 1561 GCGGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACCGCGTGTAC 1620
Qy 1627 TAGGACCCCGCAGCAGGACCTGGTGGCGGAGATCCAGAGCAGGGCGCAGCAGTGGACC 1686
Db 1621 TAGGACCCCGCAGCAGGACCTGGTGGCGGAGATCCAGAGCAGGGCGCAGCAGTGGACC 1680
```

```
Qy 1687 TACCAGATCTACGAGGAGCCCTTCAAGAACTGAAAGACCGGCAAGTACGCCAAGATGGC 1746
Db 1681 TACCAGATCTACGAGGAGCCCTTCAAGAACTGAAAGACCGGCAAGTACGCCAAGATGGC 1740
Qy 1747 ACCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGCCGTGCAGAAAGTCCCATGGAG 1806
Db 1741 ACCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGCCGTGCAGAAAGTCCCATGGAG 1800
Qy 1807 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTCGCCCATCCAGAAGAGACCTGG 1866
Db 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTCGCCCATCCAGAAGAGACCTGG 1860
Qy 1867 GAGACCTGTGTGAACCGACTACTGTGAGGCCACCTTGGATCCCGAGTGGGAGTTCTGTGAAC 1926
Db 1861 GAGACCTGTGTGAACCGACTACTGTGAGGCCACCTTGGATCCCGAGTGGGAGTTCTGTGAAC 1920
Qy 1927 ACCCCCGCCCTGGTGAAGCTGTGTGATCCAGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1986
Db 1921 ACCCCCGCCCTGGTGAAGCTGTGTGATCCAGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1980
Qy 1987 ACCTTTCTACGTGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACCTGTG 2046
Db 1981 ACCTTTCTACGTGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACCTGTG 2040
Qy 2047 ACCGACCGGGCCCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2106
Db 2041 ACCGACCGGGCCCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Qy 2107 CTGCAAGCCATTCAGCTGGGCCCTGCGAGACGCGCGCAGGAGTGAACATCGTGACCGAC 2166
Db 2101 CTGCAAGCCATTCAGCTGGGCCCTGCGAGACGCGCGCAGGAGTGAACATCGTGACCGAC 2160
Qy 2167 AGCCAGTACGCCCTGGGCGCATCATCAGGCCCGAGCCCGACAGAGCGAGCGAGCTGGTG 2226
Db 2161 AGCCAGTACGCCCTGGGCGCATCATCAGGCCCGAGCCCGACAGAGCGAGCGAGCTGGTG 2220
Qy 2227 AACCAAGATCTACGAGCAGCTGATCAAGAGAGAGTGTACTCTGAGCTGGGTGCCCGCC 2286
Db 2221 AACCAAGATCTACGAGCAGCTGATCAAGAGAGAGTGTACTCTGAGCTGGGTGCCCGCC 2280
Qy 2287 CACAAGGGCATCGGCGCAACGAGCAGATCGAAGCTGGTGAAGAGGCGATCCGCAAG 2346
Db 2281 CACAAGGGCATCGGCGCAACGAGCAGATCGAAGCTGGTGAAGAGGCGATCCGCAAG 2340
Qy 2347 GTGCTGTTCTTGGACGCGCATCGATGGCGGATCTGATCTACAGTACATGAGACGACCTG 2406
Db 2341 GTGCTGTTCTTGGACGCGCATCGATGGCGGATCTGATCTACAGTACATGAGACGACCTG 2400
Qy 2407 TAGCTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2463
Db 2401 TAGCTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457
```

## RESULT 5

ABL39960

ID ABL39960 standard; DNA; 2463 BP.

XX ABL39960;

AC ABL39960;

XX 15-MAY-2002 (first entry)

XX Synthetic construct PR975YM SEQ ID NO:31.

XX Human immunodeficiency virus type C; antigenic HIV type C protein; nef;

XX immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;

XX immunostimulant; gene therapy; gene; ds.

XX Human immunodeficiency virus; type C.

OS Synthetic.

XX W0200204493-A2.

PN

XX

XX

PD	17-JAN-2002.	QY	541	GTGCGCCAGTACGACCAAGATCTCGAGATCTGCGGCAAGAGCCATCGCACCGTG	600
XX		DB	541		
XX	05-JUL-2001; 2001WO-US021241.				
PR				GTGCGCCAGTACGACCAAGATCTCGAGATCTGCGGCAAGAGCCATCGCACCGTG	600
XX		QY	601	CTGATCGGCCCAACCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC	660
XX		DB	601		
PA	(CHIR ) CHIRON CORP.				
PA	(UYST-) UNIV STELLENBOSCH.			CTGATCGGCCCAACCCCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGC	660
XX		QY	661	ACCTCGAACTTCCCATCATCGACCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG	720
PI	Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	DB	661		
XX				ACCTTGAACTTCCCATCATCGACCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG	720
DR	WPI; 2002-154920/20.	QY	721	GACGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC	780
XX		DB	721		
PT	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in				
PT	applications including DNA immunization or generation of packaging cell				
PT	lines, particularly in gene therapy.			GACGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC	780
XX		QY	781	ATCTCGAGGAGATGAGAGAAGGGGCAAGATCACCAGATCGGCCCGGAGAACCCCTAC	840
PS	Claim 1; Fig 9; 233pp; English.	DB	781		
XX				ATCTCGAGGAGATGAGAGAAGGGGCAAGATCACCAGATCGGCCCGGAGAACCCCTAC	840
CC	The present invention describes expression cassettes comprising a	QY	841	AACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC	900
CC	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV	DB	841		
CC	type C polypeptides. The expression cassettes comprise any of the HIV				
CC	C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef				
CC	(i) have immunostimulant activity and can be used in gene therapy.				
CC	The HIV type C polynucleotides are useful in applications including DNA	QY	901	TTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCAATCCCCCAC	960
CC	immunisation, generation of packaging cell lines, and production of HIV	DB	901		
CC	Type C proteins. The polynucleotides are particularly useful in gene				
CC	therapy and DNA immunisation applications. ABL39942 to ABL40054 and	QY	961	CCGCGCGGCTGAGAGAAGAGAGCGGTGACCGTCTGGACGTGGCGGCGACTACTTC	1020
CC	ABB06204 to ABB06215 represent sequences used in the exemplification of	DB	961		
XX	the present invention				
SQ	Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;	QY	1021	AGCGTCCCTGGACGAGGACTTCCCAAGTACACCGCCTTCAACCATCCCCAGCATCAAC	1080
	Query Match 98.9%; Score 2442.2; DB 6; Length 2463;	DB	1021		
	Best Local Similarity 99.6%; Pred. No. 1.6e-294;				
	Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;	QY	1081	AACGAGACCCCGGCATCCGCTACCAAGTACAAAGTGTGCTGCCAGGGCTGGAAGGGCAGC	1140
		DB	1081		
QY	1 GTGAGCGCCACCATGGCCGAGGCCATGAGCAGGCCACACAGCCCAACATCTGTATGCAG				
DB					
QY	61 CGCAGCAACTTCAAGGGCCCAAGCGGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGC				
DB					
QY	61 CGCAGCAACTTCAAGGGCCCAAGCGGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGC				
DB					
QY	121 CACATGCGCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG				
DB					
QY	181 GGCACACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCC				
DB					
QY	181 GGCACACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCC				
DB					
QY	241 TTCCCGCCAGGCGAGGCCCGCGAGTTCCCGAGCGAGCAACCGCGCAACAGCCCCACC				
DB					
QY	241 TTCCCGCCAGGCGAGGCCCGCGAGTTCCCGAGCGAGCAACCGCGCAACAGCCCCACC				
DB					
QY	301 AGCGCGAGTGTAGGTGCGCGGCGCAACCCCGCAGCGAGCGCGCGCGCGAGCCAG				
DB					
QY	301 AGCGCGAGTGTAGGTGCGCGGCGCAACCCCGCAGCGAGCGCGCGCGAGCCAG				
DB					
QY	361 GGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCTTGTGAGCATCAAGGTG				
DB					
QY	361 GGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCTTGTGAGCATCAAGGTG				
DB					
QY	421 GCGCGCCAGATCAAGGAGGCCCTGCTGGACACCGCGCCGACGACACCGTGTGGAGGAG				
DB					
QY	421 GCGCGCCAGATCAAGGAGGCCCTGCTGGACACCGCGCCGACGACACCGTGTGGAGGAG				
DB					
QY	481 ATGAGCTGCGCGCAAGTGGAGCCCAAGATGATGGCGGCATCGCGGCTTCATCAAG				
DB					
QY	481 ATGAGCTGCGCGCAAGTGGAGCCCAAGATGATGGCGGCATCGCGGCTTCATCAAG				
DB					





QY 421 GCGGCCAGATCAAGGAGGCCCTGTCTGGACACCGCGCGCGACGACACCGTGTCTGGAGGAG 480  
 Db |||||  
 QY 421 GCGGCCAGATCAAGGAGGCCCTGTCTGGACACCGCGCGCGACGACACCGTGTCTGGAGGAG 480  
 Db |||||  
 QY 481 ATGAGCCTGCGCGCAAGTGGAGCCCAAGATGATCGGCGGCATCGCGGCTTTCATCAAG 540  
 Db |||||  
 QY 481 ATGAGCCTGCGCGCAAGTGGAGCCCAAGATGATCGGCGGCATCGCGGCTTTCATCAAG 540  
 Db |||||  
 QY 541 GTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGGCAGAGGCGCATCGGCAACCTG 600  
 Db |||||  
 QY 541 GTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGGCAGAGGCGCATCGGCAACCTG 600  
 Db |||||  
 QY 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 Db |||||  
 QY 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 Db |||||  
 QY 661 ACCCTGAACCTTCCCATCGACCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATG 720  
 Db |||||  
 QY 661 ACCCTGAACCTTCCCATCGACCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATG 720  
 Db |||||  
 QY 721 GACGCCCCAAGTGAAGCATGTGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
 Db |||||  
 QY 721 GACGCCCCAAGTGAAGCATGTGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
 Db |||||  
 QY 781 ATCTGCGAGGAGTGAAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840  
 Db |||||  
 QY 781 ATCTGCGAGGAGTGAAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840  
 Db |||||  
 QY 841 AACACCCCGCTGTTCCGCCATCAAGAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
 Db |||||  
 QY 841 AACACCCCGCTGTTCCGCCATCAAGAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
 Db |||||  
 QY 901 TTCGCGAGCTGAACAGCGCACCCAGACTTCTGGGAGGTGCGAGCTGGGCATCCCCAC 960  
 Db |||||  
 QY 901 TTCGCGAGCTGAACAGCGCACCCAGACTTCTGGGAGGTGCGAGCTGGGCATCCCCAC 960  
 Db |||||  
 QY 961 CCGCGCGCTGAAGAGAAGAGAGCGTGCACCGTCTGGACGTGGCGACGCTACTTC 1020  
 Db |||||  
 QY 961 CCGCGCGCTGAAGAGAAGAGAGCGTGCACCGTCTGGACGTGGCGACGCTACTTC 1020  
 Db |||||  
 QY 1021 AGCGTGCCCTGGACGAGGACTTCCGAAGTACACCGCTTCAACATCCCCAGCATCAAC 1080  
 Db |||||  
 QY 1021 AGCGTGCCCTGGACGAGGACTTCCGAAGTACACCGCTTCAACATCCCCAGCATCAAC 1080  
 Db |||||  
 QY 1081 AACGAGACCCCGGCATCCGCTACAGTACAAAGTACACCGCTTCAACATCCCCAGCATCAAC 1140  
 Db |||||  
 QY 1081 AACGAGACCCCGGCATCCGCTACAGTACAAAGTACACCGCTTCAACATCCCCAGCATCAAC 1140  
 Db |||||  
 QY 1141 CCAGAGATCTTCCAGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGCGCGCAACCCC 1200  
 Db |||||  
 QY 1141 CCAGAGATCTTCCAGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGCGCGCAACCCC 1200  
 Db |||||  
 QY 1201 GAGATCGTATCTACCAAGTACATGGAGCCTGTACGTGGGCGAGCGACCTGGAGATCGGC 1260  
 Db |||||  
 QY 1201 GAGATCGTATCTACCAAGTACATGGAGCCTGTACGTGGGCGAGCGACCTGGAGATCGGC 1260  
 Db |||||  
 QY 1261 CAGCACCGCGCCAGATCGAGGAGCTGCGCAAGCCTGCTGGCTGGGGCTTCAACACC 1320  
 Db |||||  
 QY 1261 CAGCACCGCGCCAGATCGAGGAGCTGCGCAAGCCTGCTGGCTGGGGCTTCAACACC 1320  
 Db |||||  
 QY 1255 CAGCACCGCGCCAGATCGAGGAGCTGCGCAAGCCTGCTGGCTGGGGCTTCAACACC 1314  
 Db |||||  
 QY 1321 CCGGACAAGAGCACCAAGAGGAGCCCTTCTCTGTGATGGCTACGAGCTGCACCCC 1380  
 Db |||||  
 QY 1321 CCGGACAAGAGCACCAAGAGGAGCCCTTCTCTGTGATGGCTACGAGCTGCACCCC 1380  
 Db |||||  
 QY 1315 CCGGACAAGAGCACCAAGAGGAGCCCTTCTCTGTGATGGCTACGAGCTGCACCCC 1374  
 Db |||||  
 QY 1381 GACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1440  
 Db |||||  
 QY 1375 GACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1434  
 Db |||||  
 QY 1441 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500  
 Db |||||  
 QY 1435 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494  
 Db |||||

RESULT 7  
 ACA03547  
 ID ACA03547 standard; DNA; 2457 BP.  
 XX

QY 1501 CGCCAGCTGTCAAGCTGTCTGCGCGCGCCAGAGGCCCTGACCGACATCTGTGCCCTTGACC 1560  
 Db |||||  
 QY 1495 CGCCAGCTGTCAAGCTGTCTGCGCGCGCCAGAGGCCCTGACCGACATCTGTGCCCTTGACC 1554  
 Db |||||  
 QY 1561 GAGGAGCCGAGCTGGAGCTGGCCGAGAACCGGAGATCTGCGGAGCCCGTGCAAGCC 1620  
 Db |||||  
 QY 1555 GAGGAGCCGAGCTGGAGCTGGCCGAGAACCGGAGATCTGCGGAGCCCGTGCAAGCC 1614  
 Db |||||  
 QY 1621 GTGTACTACGACCCCAAGCAAGCACTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAG 1680  
 Db |||||  
 QY 1615 GTGTACTACGACCCCAAGCAAGCACTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAG 1674  
 Db |||||  
 QY 1681 TGGACCTTACAGATCTTACGAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTACGCAAG 1740  
 Db |||||  
 QY 1675 TGGACCTTACAGATCTTACGAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTACGCAAG 1734  
 Db |||||  
 QY 1741 ATGCGCACCGCCCAACCAAGCACTGTGAAGCACTGACCGAGGCCCTGTCAGAAAGATCGCC 1800  
 Db |||||  
 QY 1735 ATGCGCACCGCCCAACCAAGCACTGTGAAGCACTGACCGAGGCCCTGTCAGAAAGATCGCC 1794  
 Db |||||  
 QY 1801 ATGCGAGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCCTCCAGAAAGGAG 1860  
 Db |||||  
 QY 1795 ATGCGAGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCCTCCAGAAAGGAG 1854  
 Db |||||  
 QY 1861 ACCTGGAGACCTGTGTGACCGACTACTGTGCGCAGGCCACTCTGGATCCCCAGTGGGAGTTC 1920  
 Db |||||  
 QY 1855 ACCTGGAGACCTGTGTGACCGACTACTGTGCGCAGGCCACTCTGGATCCCCAGTGGGAGTTC 1914  
 Db |||||  
 QY 1921 GTGAACACCCCGCTGTGTGAGCTGTGTGATCCAGCTGGAGAGGAGCCCATCTCGGC 1980  
 Db |||||  
 QY 1915 GTGAACACCCCGCTGTGTGAGCTGTGTGATCCAGCTGGAGAGGAGCCCATCTCGGC 1974  
 Db |||||  
 QY 1981 GCGGAGACCTTCTACGTGTGACCGCGCCCAACCGGAGACCAAGATCGCAAGGCGCGGC 2040  
 Db |||||  
 QY 1975 GCGGAGACCTTCTACGTGTGACCGCGCCCAACCGGAGACCAAGATCGCAAGGCGCGGC 2034  
 Db |||||  
 QY 2041 TAGTGACCGACCGGGCGCGGAGAAAGATCTGTGAGCTGACCGAGACCAACCAAGAAAG 2100  
 Db |||||  
 QY 2035 TAGTGACCGACCGGGCGCGGAGAAAGATCTGTGAGCTGACCGAGACCAACCAAGAAAG 2094  
 Db |||||  
 QY 2101 ACCGAGCTCAGGCCCATCCAGCTGGCCCTGCAAGGACAGCGGCGAGGAGTGAACATCGTG 2160  
 Db |||||  
 QY 2095 ACCGAGCTCAGGCCCATCCAGCTGGCCCTGCAAGGACAGCGGCGAGGAGTGAACATCGTG 2154  
 Db |||||  
 QY 2161 ACCGAGCAGCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAAGAGCGAGAGCGAG 2220  
 Db |||||  
 QY 2155 ACCGAGCAGCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAAGAGCGAGAGCGAG 2214  
 Db |||||  
 QY 2221 CTGGTGAACCAAGATCATCGAGCAGCTGTATCAAGAAAGGAGAAAGTGTACTCTGAGCTGGGTG 2280  
 Db |||||  
 QY 2215 CTGGTGAACCAAGATCATCGAGCAGCTGTATCAAGAAAGGAGAAAGTGTACTCTGAGCTGGGTG 2274  
 Db |||||  
 QY 2281 CCGGCCCAACAGGGCATCGGCGCAACGAGCAGATCGAAGCTGTGAGCAAGGGGCATC 2340  
 Db |||||  
 QY 2275 CCGGCCCAACAGGGGCATCGGCGCAACGAGCAGATCGAAGCTGTGAGCAAGGGGCATC 2334  
 Db |||||  
 QY 2341 CGCAAGTGTCTGTCTCTGGACCGCATCTGATGGCGCATCTGTGATCTACCAAGTACATGGAC 2400  
 Db |||||  
 QY 2335 CGCAAGTGTCTGTCTCTGGACCGCATCTGATGGCGCATCTGTGATCTACCAAGTACATGGAC 2394  
 Db |||||  
 QY 2401 GACCTGTAGTGGGCGAGCGGCCCTTAGATCGATTAAAGCTTCCCGGGGCTTAGCACC 2460  
 Db |||||  
 QY 2395 GACCTGTAGTGGGCGAGCGGCCCTTAGATCGATTAAAGCTTCCCGGGGCTTAGCACC 2454  
 Db |||||  
 QY 2461 GGTGAATTC 2469  
 Db |||||  
 QY 2455 GGTGAATTC 2463  
 Db |||||



Db 1315 CCGCAAGAAGCACACAGAGGAGCCCCCTTCTCTGTGGATGGCTACGAGCTGCACCCC 1374

Qy 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTCCCGAGAGGAGAGCTGGACCCGTGAACGAC 1440

Db 1375 GACAAGTGGACCGTGCAGCCCATCGAGCTCCCGAGAGGAGAGCTGGACCCGTGAACGAC 1434

Qy 1441 ATCCAGAAGCTGTGGGCAAGCTGAACTGGCGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Db 1435 ATCCAGAAGCTGTGGGCAAGCTGAACTGGCGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494

Qy 1501 CGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCCCTGACGACATCTGTGCCCTGACC 1560

Db 1495 CGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCCCTGACGACATCTGTGCCCTGACC 1554

Qy 1561 GAGGAGCCGAGCTGAGCTGGCCGCGAGAACCGGAGATCTGCGCGAGCCCGTGCAGGC 1620

Db 1555 GAGGAGCCGAGCTGAGCTGGCCGCGAGAACCGGAGATCTGCGCGAGCCCGTGCAGGC 1614

Qy 1621 GTCTACTACGACCCACAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGCGCACGACCAG 1680

Db 1615 GTGTACTACGACCCACAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGCGCACGACCAG 1674

Qy 1681 TGGACCTTACCAGATCTACCAAGGAGCCCTTCAAGAACTGAAGACCGCGCAAGTACGCCAAG 1740

Db 1675 TGGACCTTACCAGATCTACCAAGGAGCCCTTCAAGAACTGAAGACCGCGCAAGTACGCCAAG 1734

Qy 1741 ATCGGACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGGCCGTGCAGAGATCGCC 1800

Db 1735 ATCGGACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGGCCGTGCAGAGATCGCC 1794

Qy 1801 ATGGAGAGCATCTGTGCTGGGCGAGACCCCAAGTTCGCGCTGCCCATCCAGAAGGAG 1860

Db 1795 ATGGAGAGCATCTGTGCTGGGCGAGACCCCAAGTTCGCGCTGCCCATCCAGAAGGAG 1854

Qy 1861 ACCTGGGAGACCTGTGGGACCGCACTACTGGCAGGCCACCTGGATCCCCGAGTGGGATTC 1920

Db 1855 ACCTGGGAGACCTGTGGGACCGCACTACTGGCAGGCCACCTGGATCCCCGAGTGGGATTC 1914

Qy 1921 GTGAACACCCCCCTGTGTGAAGCTGTGTATCAGCTGAGAGAGGCCCATCATCGGC 1980

Db 1915 GTGAACACCCCCCTGTGTGAAGCTGTGTATCAGCTGAGAGAGGCCCATCATCGGC 1974

Qy 1981 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2040

Db 1975 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2034

Qy 2041 TAGCTGACCGAGCGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAG 2100

Db 2035 TAGCTGACCGAGCGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAG 2094

Qy 2101 ACCGAGCTGACGGCCATCCAGCTGGCCCTGCAGGACGCGCGAGCGGTGAACATCGTG 2160

Db 2095 ACCGAGCTGACGGCCATCCAGCTGGCCCTGCAGGACGCGCGAGCGGTGAACATCGTG 2154

Qy 2161 ACCGAGCGCAGTAGCGCCCTGGGCATCATCCAGGCCCGAGCCCGAGAGCGAGCGAG 2220

Db 2155 ACCGAGCGCAGTAGCGCCCTGGGCATCATCCAGGCCCGAGCCCGAGAGCGAGCGAG 2214

Qy 2221 CTGGTGAACCCAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTGAGCTGGGTG 2280

Db 2215 CTGGTGAACCCAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTGAGCTGGGTG 2274

Qy 2281 CCCGCCACAAAGGCGATCGCGGCAAGCAGATCGACAAGCTGGTGGAGCAAGGGCGATC 2340

Db 2275 CCCGCCACAAAGGCGATCGCGGCAAGCAGATCGACAAGCTGGTGGAGCAAGGGCGATC 2334

Qy 2341 CGCAAGTGTCTGTTCCTGGAGCGCATCGATGGCGGCATCTGTATCTACCAAGTACATGGAC 2400

Db 2335 CGCAAGTGTCTGTTCCTGGAGCGCATCGATGGCGGCATCTGTATCTACCAAGTACATGGAC 2394

Qy 2401 GACCTGTACCTGGGCGAGCGCGCCCTTAGGATCGATTAAGCTTCCCGGGCTTAGCAC 2460

Db 2395 GACCTGTACCTGGGCGAGCGCGCCCTTAGGATCGATTAAGCTTCCCGGGCTTAGCAC 2454

Qy 2461 GGT 2463

Db 2455 GGT 2457

RESULT 8

ADCL13265

ID ADCL13265 standard; DNA; 2457 BP.

XX

AC ADCL13265;

XX

DT 18-DEC-2003 (first entry)

XX

DE DNA of HIV construct p2Pol-opt-YM\_C SEQ ID NO 44.

XX

KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;

KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX

OS Human immunodeficiency virus.

XX

PN WO2003004620-A2.

XX

PD 16-JAN-2003.

XX

PF 05-JUL-2002; 2002WO-US021420.

XX

PR 05-JUL-2001; 2001US-0303192P.

PR 31-AUG-2001; 2001US-0316860P.

PR 16-JAN-2002; 2002US-0349871P.

XX

(CHIR ) CHIRON CORP.

(UYST-) UNIV STELLENBOSCH.

Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

WPI; 2003-221593/21.

XX New expression cassette comprising a polynucleotide sequence encoding a

PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

PT Prot, or Rev polypeptide, useful for immunization, or generating

PT packaging cell lines.

XX

PS Disclosure; Fig 41; 301pp; English.

XX

CC The invention relates to a novel expression cassette comprising a

CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,

CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel

CC expression cassette can be used to treat HIV type C by gene therapy or

CC used in the development of a vaccine. The gene delivery vector is

CC administered intramuscularly, intramucosally, intranasally,

CC subcutaneously, intradermally, transdermally, intravaginally,

CC intrarectally, orally or intravenously. The expression cassette is useful

CC for immunisation, generating packaging cell lines and producing HIV

CC polypeptides. This polynucleotide sequence represents the DNA of an HIV

CC Type C related sequence of the invention.

XX

Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.6%; Score 2434.6; DB 10; Length 2457;

Best Local Similarity 99.6%; Pred. No. 1.4e-293;

Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 1 GTGAGCGCCACCATGCGCGGCGCATGAGCCAGCCAGCCAGCCACATCTCTGATCGAG 60

Db 1 GTGAGCGCCACCATGCGCGGCGCATGAGCCAGCCAGCCAGCCACATCTCTGATCGAG 60

Qy 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGC 120

Db 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGC 120

Qy 121 CACATCGCCCGCAACTCGCGCGCCCCCGCCAGAGGGCTGTGGAGTCCGGCAAGAG 180

Db 121 CACATCGCCGCACTCGCGGCCCCCGCAAGAGGGCTGCTGAAAGTGGGCAAGGAG 180  
Qy 181 GGCACACAGATGAAGGATGTCACAGCGCCGCAAGGCCAACTTCTTCGCGAGGAACTTGGCC 240  
Db 181 GGCACACAGATGAAGGATGTCACAGCGCCGCAAGGCCAACTTCTTCGCGAGGAACTTGGCC 240  
Qy 241 TTCCCCAGGSCAAGGCCCGGAGTTCCCGACGAGCAGAAACCGCGCCAAACAGCCCCACC 300  
Db 241 TTCCCCAGGSCAAGGCCCGGAGTTCCCGACGAGCAGAAACCGCGCCAAACAGCCCCACC 300  
Qy 301 AGCCGCGAGCTGCAAGTGGCGGCAACAAACCCCGACGAGCGCGCGCCGAGCGCCAG 360  
Db 301 AGCCGCGAGCTGCAAGTGGCGGCAACAAACCCCGACGAGCGCGCGCCGAGCGCCAG 360  
Qy 361 GGCACCTTGAACTTCCCCAGATCACTCTGTGGCAGCGCCCTGCTGGTGAAGCATCAAGGTG 420  
Db 361 GGCACCTTGAACTTCCCCAGATCACTCTGTGGCAGCGCCCTGCTGGTGAAGCATCAAGGTG 420  
Qy 421 GCGCGCAGATCAAGGAGGCTCTGTGACACCGCGCGGACGACACCGTCTGGAGGAG 480  
Db 421 GCGCGCAGATCAAGGAGGCTCTGTGACACCGCGCGGACGACACCGTCTGGAGGAG 480  
Qy 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG 540  
Db 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG 540  
Qy 541 GTGCGCCAGTACGACCAAGATCCTGATCGAGATCTCGCGCAAGAGGCCATCGGACCGTG 600  
Db 541 GTGCGCCAGTACGACCAAGATCCTGATCGAGATCTCGCGCAAGAGGCCATCGGACCGTG 600  
Qy 601 CTGATCGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGAACCCAGCTGGGTGC 660  
Db 601 CTGATCGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGAACCCAGCTGGGTGC 660  
Qy 661 ACCCTGAACCTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG 720  
Db 661 ACCCTGAACCTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG 720  
Qy 721 GACGCGCCCAAGGTGAAGTGAAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTTACCGCC 780  
Db 721 GACGCGCCCAAGGTGAAGTGAAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTTACCGCC 780  
Qy 781 ATCTGCGAGGAGATGGAAGAGGCGGCAAGATCACCAAGATCGGCGCCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGGAAGAGGCGGCAAGATCACCAAGATCGGCGCCGAGAACCCCTAC 840  
Qy 841 AACACCCCGTGTTCGCCATCAAGAAGAGACAGCAACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGTGTTCGCCATCAAGAAGAGACAGCAACCAAGTGGCGCAAGCTGGTGGAC 900  
Qy 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCAATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCAATCCCCAC 960  
Qy 961 CCGCGCGCTGAAGAAGAGAGCGTGACCGTGTGACGCTGGGCGAGCGCTACTTTC 1020  
Db 961 CCGCGCGCTGAAGAAGAGAGCGTGACCGTGTGACGCTGGGCGAGCGCTACTTTC 1020  
Qy 1021 AGCGTGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGGCATCCGCTACAGTACAAACGCTGTGCCCGCAGGGCTGGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGGCATCCGCTACAGTACAAACGCTGTGCCCGCAGGGCTGGAAGGGCAGC 1140  
Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGTGAGGCCCTTTCGCGCCCGCAACCCC 1200  
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGTGAGGCCCTTTCGCGCCCGCAACCCC 1200  
Qy 1201 GAGATCGTATCTACCACTACATGACGACTGTACGTGGCGAGGACCTGGAGATCGGC 1260  
Db 1201 GAGATCGTATCTACCA-----GGCCCCCTGTACGTGGCGAGCGACCTGGAGATCGGC 1254

Qy 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTTGTGCTGCGTGGGGCTTCAACCAC 1320  
Db 1255 CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTTGTGCTGCGTGGGGCTTCAACCAC 1314  
Qy 1321 CCGCACAGAGCACCAGAGAGCCCTTCTGTGTGGATGGGCTACGAGCTGCACGCC 1380  
Db 1315 CCGCACAGAGCACCAGAGAGCCCTTCTGTGTGGATGGGCTACGAGCTGCACGCC 1374  
Qy 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAAGAGGAGCTGCAACCTGAAACGAC 1440  
Db 1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAAGAGGAGCTGCAACCTGAAACGAC 1434  
Qy 1441 ATCCAGAACTGTGTGGGCAAGCTGAATGGGCGCAGCCAGATCTACCCCGGATCAAGGTG 1500  
Db 1435 ATCCAGAACTGTGTGGGCAAGCTGAATGGGCGCAGCCAGATCTACCCCGGATCAAGGTG 1494  
Qy 1501 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCAAGGCCCTGAGCGACATCGTGCCTGACC 1560  
Db 1495 CGCCAGCTGTGCAAGCTGCTGCGCGGCGCAAGGCCCTGAGCGACATCGTGCCTGACC 1554  
Qy 1561 GAGGAGGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGCC 1620  
Db 1555 GAGGAGGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGCC 1614  
Qy 1621 GTGTACTACGACCCAGCAAGGACTGTGTGGCGCAGATCTCCAGAAAGAGGCGCAACGAC 1680  
Db 1615 GTGTACTACGACCCAGCAAGGACTGTGTGGCGCAGATCTCCAGAAAGAGGCGCAACGAC 1674  
Qy 1681 TGGACCTTACCAGATCTACAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAG 1740  
Db 1675 TGGACCTTACCAGATCTTACAGGAGCCCTTCAAGAACTGGAAGACCGGCAAGTACGCCAAG 1734  
Qy 1741 ATGCGCACCGCCACACCAAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCC 1800  
Db 1735 ATGCGCACCGCCACACCAAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCC 1794  
Qy 1801 ATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCCCTGCCCATCCAGAAAGAG 1860  
Db 1795 ATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCCCTGCCCATCCAGAAAGAG 1854  
Qy 1861 ACCTGGAGAGCTTGGTGGACCGACTACTGCGAGCGCCACCTGGATCCCCAGTGGGAGTTC 1920  
Db 1855 ACCTGGAGAGCTTGGTGGACCGACTACTGCGAGCGCCACCTGGATCCCCAGTGGGAGTTC 1914  
Qy 1921 GTGAACACCCCGCCCTGGTGAAGCTGTGTGTACAGCTGGAGAAAGAGGCCCATCATCGGC 1980  
Db 1915 GTGAACACCCCGCCCTGGTGAAGCTGTGTGTACAGCTGGAGAAAGAGGCCCATCATCGGC 1974  
Qy 1981 GCCGAGACCTTCTACGTGGACGCGCCGCAACCCGCGAGACCAAGATCGGCAAGGCCCGGC 2040  
Db 1975 GCCGAGACCTTCTACGTGGACGCGCGCCGCAACCCGCGAGACCAAGATCGGCAAGGCCCGGC 2034  
Qy 2041 TACGTGACCGACCGGGCGCGCAGAAATCGTGAAGCTTACCGAGACCGACCAACCGAGAG 2100  
Db 2035 TACGTGACCGACCGGGCGCGCAGAAATCGTGAAGCTTACCGAGACCGACCAACCGAGAG 2094  
Qy 2101 ACCGAGCTGCAAGGCGCATTCAGCTGCCCTGCAAGACAGCGCAGCGAGGTGAAATCTGTG 2160  
Db 2095 ACCGAGCTGCAAGGCGCATTCAGCTGCCCTGCAAGACAGCGCAGCGAGGTGAAATCTGTG 2154  
Qy 2161 ACCGAGACCGGATACGCCCTTGGGCAATCATCCAGGCCCGAGCCCGCAAGAGCGAGCGAG 2220  
Db 2155 ACCGAGACCGGATACGCCCTTGGGCAATCATCCAGGCCCGAGCCCGCAAGAGCGAGCGAG 2214  
Qy 2221 CTGTGAAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTTGATCTAGCTGGGTG 2280  
Db 2215 CTGTGAAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTTGATCTAGCTGGGTG 2274  
Qy 2281 CCGCCCCACAGGGGCATCGCGGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGGCATC 2340  
Db 2275 CCGCCCCACAGGGGCATCGCGGCAACGAGCAGATCGACAGCTGGTGGAGCAAGGGCATC 2334

QY	2341	CGAAGTGTCTTCTCTGGAGCGCATCGATGGCGCATCGTATCTACCAAGTACATGGAC	2400
DB	2335	CGCAAGGTGTCTTCTCTGGAGCGCATCGATGGCGCATCGTATCTACCAAGTACATGGAC	2394
QY	2401	GACCTGTACGTGGCGAGCGCGCCCTTAGATCGATTTAAAGCTTCCCGGGCTAGCACC	2460
DB	2395	GACCTGTACGTGGCGAGCGCGCCCTTAGATCGATTTAAAGCTTCCCGGGCTAGCACC	2454
QY	2461	GGT 2463	
DB	2455	GGT 2457	
RESULT 9			
ID	ABL39961	standard; DNA; 2457 BP.	
XX	AC	ABL39961;	
XX	DT	15-MAY-2002 (first entry)	
XX	XX	Synthetic construct PR975YMM SEQ ID NO:32.	
XX	DE	Human immunodeficiency virus type C; antigenic HIV type C protein;	
KW	KW	immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;	
KW	KW	immunostimulant; gene therapy; gene; ds.	
XX	XX	Human immunodeficiency virus; type C.	
OS	OS	Synthetic.	
XX	XX	WO200204493-A2.	
XX	PN	17-JAN-2002.	
XX	PD	05-JUL-2001; 2001WO-US021241.	
XX	PF	05-JUL-2000; 2000US-00610313.	
XX	PR	(CHIR ) CHIRON CORP.	
XX	XX	(UYST-) UNIV STELLENBOSCH.	
PA	PA	Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	
XX	PI	WPI; 2002-154920/20.	
XX	DR	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in	
XX	XX	applications including DNA immunization or generation of packaging cell	
PT	PT	lines, particularly in gene therapy.	
XX	XX	Claim 1; Fig 10; 233pp; English.	
XX	XX	The present invention describes expression cassettes comprising a	
CC	CC	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV	
CC	CC	type C polypeptides. The expression cassettes comprise any of the HIV	
CC	CC	type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef	
CC	CC	(I). (I) have immunostimulant activity and can be used in gene therapy.	
CC	CC	The HIV type C polynucleotides are useful in applications including DNA	
CC	CC	immunisation, generation of packaging cell lines, and production of HIV	
CC	CC	type C proteins. The polynucleotides are particularly useful in gene	
CC	CC	therapy and DNA immunisation applications. ABL39942 to ABL40054 and	
CC	CC	ABB06204 to ABB06215 represent sequences used in the exemplification of	
CC	CC	the present invention	
XX	XX	Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;	
SQ	SQ	Query Match 97.8%; Score 2415.4; DB 6; Length 2457;	
		Best Local Similarity 99.3%; Pred. No. 3.3e-291;	
		Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;	
QY	1	GTGAGCGCCACCATGGCCGAGGCGCATGAGCAGGCCACCGAGCGCCCAACATCTCTGATGCGAG	60
DB	1	GTGAGCGCCACCATGGCCGAGGCGCATGAGCAGGCCACCGAGCGCCCAACATCTCTGATGCGAG	60

QY	61	CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGC	120
DB	61	CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGC	120
QY	121	CACATCGCGCGCAACTGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGCAGGCAAGGAG	180
DB	121	CACATCGCGCGCAACTGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGCAGGCAAGGAG	180
QY	181	GGCCACCAGATGAAGGACTGCAACCGAGCGCCAGGCCCAACTTCTTCGCGAGGACCTTGGCC	240
DB	181	GGCCACCAGATGAAGGACTGCAACCGAGCGCCAGGCCCAACTTCTTCGCGAGGACCTTGGCC	240
QY	241	TTCCCCCAGGCGAAGGCCCGCGAGTTTCCCCAGCGAGCAGAACCGCCCAACAGCCCCACC	300
DB	241	TTCCCCCAGGCGAAGGCCCGCGAGTTTCCCCAGCGAGCAGAACCGCCCAACAGCCCCACC	300
QY	301	AGCGCGAGCTGCGAGTGCAGGCGCGCAAAACCCCGAGCGAGGCCCGCGCGAGCGCGAG	360
DB	301	AGCGCGAGCTGCGAGTGCAGGCGCGCAAAACCCCGAGCGAGGCCCGCGCGAGCGCGAG	360
QY	361	GGCACCCTGAACCTTCCCCCAGATCACCTGTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG	420
DB	361	GGCACCCTGAACCTTCCCCCAGATCACCTGTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG	420
QY	421	GGCGGCGAGATCAAGGAGGCGCCCTGTGTGAGCACCGCGCGCGAGCACACCGTGTGAGGAG	480
DB	421	GGCGGCGAGATCAAGGAGGCGCCCTGTGTGAGCACCGCGCGCGAGCACACCGTGTGAGGAG	480
QY	481	ATGAGCCTTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAG	540
DB	481	ATGAGCCTTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAG	540
QY	541	GTGCGCGAGTACGACCAAGATCCTGTATCGAGATCTCGGCGAAGAGGCCATCGCACCGTG	600
DB	541	GTGCGCGAGTACGACCAAGATCCTGTATCGAGATCTCGGCGAAGAGGCCATCGCACCGTG	600
QY	601	CTGATCGGCGCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
DB	601	CTGATCGGCGCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
QY	661	ACCTGAACTTCCCATCGACCCCATCGAGACCGTGCCTGAGCTGAAGTGAAGCCCGGATG	720
DB	661	ACCTGAACTTCCCATCGACCCCATCGAGACCGTGCCTGAGCTGAAGTGAAGCCCGGATG	720
QY	721	GACGCGCCCAAGGTGAAGCAAGTGGCCCTGACGAGGAGAAAGATCAAGGCCCTTGACCGCC	780
DB	721	GACGCGCCCAAGGTGAAGCAAGTGGCCCTGACGAGGAGAAAGATCAAGGCCCTTGACCGCC	780
QY	781	ATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCGCCCGAGAACCCCTAC	840
DB	781	ATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCGCCCGAGAACCCCTAC	840
QY	841	AACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAC	900
DB	841	AACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAC	900
QY	901	TTCCGCGAGTGAACCAAGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCATCCCCCAC	960
DB	901	TTCCGCGAGTGAACCAAGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCATCCCCCAC	960
QY	961	CCCGCGCGCTCAAGAAAGAGAGAGCGTGCACCGTGTGACGCTGGGCGCAGCGCTACTTC	1020
DB	961	CCCGCGCGCTCAAGAAAGAGAGAGCGTGCACCGTGTGACGCTGGGCGCAGCGCTACTTC	1020
QY	1021	AGCGTGCCTTGAGCGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC	1080
DB	1021	AGCGTGCCTTGAGCGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC	1080
QY	1081	AACGAGACCCCGGCGATCCGCTACCAAGTACCAAGTGTGCGCGCGAGGCGGAGGCGAGC	1140
DB	1081	AACGAGACCCCGGCGATCCGCTACCAAGTACCAAGTGTGCGCGCGAGGCGGAGGCGAGC	1140
QY	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGGCCCTTTCGCGCGCCGCAACCC	1200



Db 1141 CCCAGCATCTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCAACCCC 1200  
Qy 1201 GAGATCGTGATCTTACCAAGATGATGAGACCTGTAGTGGGCGAGCGACTTGGAGATCGGC 1260  
Db 1201 GAGATCGTGATCTTACCA-----GGCCCCCTGTAGTGGGCGAGCGACTTGGAGATCGGC 1254  
Qy 1261 CAGACCGCGCGCAGATGAGAGCTGGCAGAGCTGGCAAGCACTGTGCTGGTGGGGCTTACACACC 1320  
Db 1255 CAGACCGCGCGCAGATGAGAGCTGGCAGAGCTGGCAAGCACTGTGCTGGTGGGGCTTACACACC 1314  
Qy 1321 CCCGACAAAGAGCACCAGAGAGCCCTTCTGTGTGATGGGCTACGAGCTGCACCCC 1380  
Db 1315 CCCGACAAAGAGCACCAGAGAGCCCTTCTGTGTGATGGGCTTCTGTGCCCCAT-----CGAGCTGCACCCC 1368  
Qy 1381 GACAAGTGGACCGGTGACGCCATCGAGCTGGCCCGAGAGAGAGCTGGACCGTGAACGAC 1440  
Db 1369 GACAAGTGGACCGGTGACGCCATCGAGCTGGCCCGAGAGAGAGCTGGACCGTGAACGAC 1428  
Qy 1441 ATCCAGAAGCTGGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCGATCAAGGTG 1500  
Db 1429 ATCCAGAAGCTGGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCGATCAAGGTG 1488  
Qy 1501 CGCCAGCTGTCAAGCTGCTGGCGGGCCCAAGGCCCTTGACCGGACATGTCGCCCTGACC 1560  
Db 1489 CGCCAGCTGTCAAGCTGCTGGCGGGCCCAAGGCCCTTGACCGGACATGTCGCCCTGACC 1548  
Qy 1561 GAGGAGGCCAGCTGGAGCTGGCGCGAGAAACCGCGAGATCTTCCGCGAGCCCGTGCACGGC 1620  
Db 1549 GAGGAGGCCAGCTGGAGCTGGCGCGAGAAACCGCGAGATCTTCCGCGAGCCCGTGCACGGC 1608  
Qy 1621 GTGTACTACGACCCCGCAGAGGAACTGGTGGCGCGAGATCTTCCGCGAGCCCGTGCACGGC 1680  
Db 1609 GTGTACTACGACCCCGCAGAGGAACTGGTGGCGCGAGATCTTCCGCGAGCCCGTGCACGGC 1668  
Qy 1681 TGGACCTACCAAGCTTACCAAGGCGCTTCAAGAACCTGAGACCGCGCAAGTACGCCAAG 1740  
Db 1669 TGGACCTACCAAGCTTACCAAGGCGCTTCAAGAACCTGAGACCGCGCAAGTACGCCAAG 1728  
Qy 1741 ATGCGCACCGCCACACCAAGCTGTAAGCAGCTGACCGAGGCGCTGCAAGAGATCGCC 1800  
Db 1729 ATGCGCACCGCCACACCAAGCTGTAAGCAGCTGACCGAGGCGCTGCAAGAGATCGCC 1788  
Qy 1801 ATGAGAGCAGCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAG 1860  
Db 1789 ATGAGAGCAGCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAG 1848  
Qy 1861 ACCTGGGAGACCTGTGACCGGACTACTGGCAGGCGCACCTGGATCCCGGATGGAGTTC 1920  
Db 1849 ACCTGGGAGACCTGTGACCGGACTACTGGCAGGCGCACCTGGATCCCGGATGGAGTTC 1908  
Qy 1921 GTGAACACCCCGCCCTGGTGAAGCTGTGTACCAAGCTGGAAGAGGAGCCCATCATCGGC 1980  
Db 1909 GTGAACACCCCGCCCTGGTGAAGCTGTGTACCAAGCTGGAAGAGGAGCCCATCATCGGC 1968  
Qy 1981 GCCGAGACCTTCTACGTGGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGC 2040  
Db 1969 GCCGAGACCTTCTACGTGGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGC 2028  
Qy 2041 TACGTGACCGACCGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAG 2100  
Db 2029 TACGTGACCGACCGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAG 2088  
Qy 2101 ACCGAGCTGAGGCGCATTCAGCTGGCCTGTGAGGACAGCGGACGAGGTGAACATCGTG 2160  
Db 2089 ACCGAGCTGAGGCGCATTCAGCTGGCCTGTGAGGACAGCGGACGAGGTGAACATCGTG 2148  
Qy 2161 ACCGAGCGCAGTACGCGCTGGGCGATCATCAGGCCCGACCGCGCAGAGCGAGCGAG 2220  
Db 2149 ACCGAGCGCAGTACGCGCTGGGCGATCATCAGGCCCGACCGCGCAGAGCGAGCGAG 2208  
Qy 2221 CTGTGAAACCGAGTATCATGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2280

Db 2209 CTGTGAAACCGAGTATCATGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2268  
Qy 2281 CCCGCCCAACAAAGGCGATCGCGGCAACGAGCAGATCGAACAAGCTGGTGGAGCAAGGCGATC 2340  
Db 2269 CCCGCCCAACAAAGGCGATCGCGGCAACGAGCAGATCGAACAAGCTGGTGGAGCAAGGCGATC 2328  
Qy 2341 CGCAAGGTGCTGTCTCTGGACGGCGATCGATCGCGGCGATCGTGTATCTACAGTACATGGAC 2400  
Db 2329 CGCAAGGTGCTGTCTCTGGACGGCGATCGATCGCGGCGATCGTGTATCTACAGTACATGGAC 2388  
Qy 2401 GACCTGTACGTGGCGCGCGCCCTAGGATCGATTAAAAGCTTCCCGGGCTAGCACC 2460  
Db 2389 GACCTGTACGTGGCGCGCGCCCTAGGATCGATTAAAAGCTTCCCGGGCTAGCACC 2448  
Qy 2461 GGTGAATTC 2469  
Db 2449 GGTGAATTC 2457  
RESULT 10  
ADM73766  
ID ADM73766 standard; DNA; 2457 BP.  
XX  
AC ADM73766;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE HIV-1 polynucleotide #9.  
XX  
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
KW HIV type C protein; immunostimulant.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN US2003223961-A1.  
PD 04-DEC-2003.  
XX  
PF 05-JUL-2001; 2001US-00899575.  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (MEGE/) MEGEDE J Z.  
PA (BARN/) BARNETT S W.  
PA (ENGE/) ENGELBRECHT S.  
PA (RENS/) RENSBERG E J V.  
XX  
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;  
XX  
DR WPI; 2004-060515/06.  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding an  
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
PT immunization, generating of packaging cell lines or in producing HIV Type  
PT C proteins.  
XX  
PS Claim 1; SEQ ID NO 32; 160pp; English.  
XX  
CC The invention relates to an expression cassette comprising a  
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
CC also relates to a recombinant expression system for use in a host cell  
CC comprising an expression cassette, where the polynucleotide sequence  
CC further comprises control elements capable of driving expression in the  
CC selected host cell, a cell comprising an expression cassette where the  
CC polynucleotide sequence further comprises control elements compatible  
CC with the expression in the cell and a composition for generating an  
CC immunological response, comprising an expression cassette. The expression  
CC cassette and the methods of the invention are useful in eliciting an  
CC immune response, in DNA immunisation, in generation of packaging cell  
CC lines and in producing HIV Type C proteins. This sequence represents an  
CC HIV-1 polynucleotide of the invention.  
XX  
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match	97.8%;	Score 2415.4;	DB 12;	Length 2457;
Best Local Similarity	99.3%;	Pred. No. 3.3e-291;		
Matches 2451; Conservative	0;	Mismatches 6;	Indels 12;	Gaps 2;

Qy	1	GTGAGCCCAACCATGGCCGAGGCCATCAGCAGGCGCCACAGCGCCACACAGCGCAACATCTCTGATCGAG	60
Db	1	GTGAGCCCAACCATGGCCGAGGCCATCAGCAGGCGCCACAGCGCCACACATCTCTGATCGAG	60
Qy	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAA	120
Db	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAA	120
Qy	121	CACATCGCCCGCAACTGTCGGGCCCGCCCGCAAGNAGGCTGCTGGAGTGGCGCAAGGAG	180
Db	121	CACATCGCCCGCAACTGTCGGGCCCGCCCGCAAGNAGGCTGCTGGAGTGGCGCAAGGAG	180
Qy	181	GGCCACCAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTCGGCC	240
Db	181	GGCCACCAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTCGGCC	240
Qy	241	TTTCCCCAGGGCAAGGCCCGCGAGTTTCCCGAGCGAGCAGAAACCGCGCCAA	300
Db	241	TTTCCCCAGGGCAAGGCCCGCGAGTTTCCCGAGCGAGCAGAAACCGCGCCAA	300
Qy	301	AGCGCGAGCTGCAGGTGCGCGGCGCAACCCCGCAGCGAGGCGCGCGCAGCGCCAG	360
Db	301	AGCGCGAGCTGCAGGTGCGCGGCGCAACCCCGCAGCGAGGCGCGCGCAGCGCCAG	360
Qy	361	GGCACCTGTAACTTTCCTCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG	420
Db	361	GGCACCTGTAACTTTCCTCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG	420
Qy	421	GGGGCCAGATCAAGAGGCGCTGCTGGACACCGCGCCCGACGACACCGTCTGGAGGAG	480
Db	421	GGGGCCAGATCAAGAGGCGCTGCTGGACACCGCGCCCGACGACACCGTCTGGAGGAG	480
Qy	481	ATGAGCTGCGCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCAATCAAG	540
Db	481	ATGAGCTGCGCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCAATCAAG	540
Qy	541	GTGGCCAGTACGACGAGATCTGATCGAGATCTGCGGCGAAGAGGCCATCGGCACCGTG	600
Db	541	GTGGCCAGTACGACGAGATCTGATCGAGATCTGCGGCGAAGAGGCCATCGGCACCGTG	600
Qy	601	CTGATCGGCGCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGTGC	660
Db	601	CTGATCGGCGCCACCCCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGTGC	660
Qy	661	ACCTGAACTTCCCATCAGCCCCCATCGACACCGTGTCCCGTGAAGCTGAAGCCCGGATG	720
Db	661	ACCTGAACTTCCCATCAGCCCCCATCGACACCGTGTCCCGTGAAGCTGAAGCCCGGATG	720
Qy	721	GACGGCCCCAAGGTGAAGCAGTGGCCCTGACGAGGAGAGATCAAGGCCCTGACCGCC	780
Db	721	GACGGCCCCAAGGTGAAGCAGTGGCCCTGACGAGGAGAGATCAAGGCCCTGACCGCC	780
Qy	781	ATCTCGAGGAGATGGAAGAAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC	840
Db	781	ATCTCGAGGAGATGGAAGAAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC	840
Qy	841	AACACCCCGTGTTCGCATCAGAAAGAGCAGCAGCAACGATGGCGCGCAGCTGGTGAC	900
Db	841	AACACCCCGTGTTCGCATCAGAAAGAGCAGCAGCAACGATGGCGCGCAGCTGGTGAC	900
Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTCGGAGGTGACGCTGGGCATCCCCAC	960
Db	901	TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTCGGAGGTGACGCTGGGCATCCCCAC	960
Qy	961	CCCGCCGCTGAAGAAGAAGAGCGTGCACGCTGTGACGCTGGCGCAGCCTACTTC	1020
Db	961	CCCGCCGCTGAAGAAGAAGAGCGTGCACGCTGTGACGCTGGCGCAGCCTACTTC	1020

Qy	1021	AGCGTCCCTTGACGAGGACTTCCGCAAGTACACGCGCTTCAACCTCCAGCATCCAGCATCAAC	1081
Db	1021	AGCGTCCCTTGGACGAGGACTTCCGCAAGTACACGCGCTTCAACCTCCAGCATCAAC	1080
Qy	1081	AAAGAGACCCCGCATCCGCTACCAAGTACAAAGTCTGCGCCCAAGGCTGGAAGGGGAGC	1140
Db	1081	AAAGAGACCCCGCATCCGCTACCAAGTACAAAGTCTGCGCCCAAGGCTGGAAGGGGAGC	1140
Qy	1141	CCAGCATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCC	1200
Db	1141	CCAGCATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCC	1200
Qy	1201	GAGATCTGTATCTACCAAGTACATGGACGACCTGTATCGTGGGCGAGCATCTGGAGATCGCG	1260
Db	1201	GAGATCTGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGCATCTGGAGATCGCG	1254
Qy	1261	CAGCACCGCGCAGATCGAGGAGCTGCGCAAGCATCTGTCTGCGCTGGGCGCTTCAACACC	1320
Db	1255	CAGCACCGCGCCCAAGATCTGAGGAGCTGCGCAAGCATCTGTCTGCGCTGGGCGCTTCAACACC	1314
Qy	1321	CCCGACAAGACACCAAGAGGAGCCCCCTCTCTGTGTAGTGGCTACGAGCTGCAACCCC	1380
Db	1315	CCCGACAAGACACCAAGAGGAGCCCCCTCTCTGTGCCAT-----CGAGCTGCACCCC	1368
Qy	1381	GACAAGTGGACCGGTGCAGCCCATCGAGCTGCCCGAAGAGAGAGCTGGACCGTGAACGAC	1440
Db	1369	GACAAGTGGACCGGTGCAGCCCATCGAGCTGCCCGAAGAGAGAGCTGGACCGTGAACGAC	1428
Qy	1441	ATCCAGAGCTGGTGGCAGCTGMACTGGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1500
Db	1429	ATCCAGAGCTGGTGGCAGCTGMACTGGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1488
Qy	1501	CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTCGACGACATCTGTGCCCTTGACC	1560
Db	1489	CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTCGACGACATCTGTGCCCTTGACC	1548
Qy	1561	GAGGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTGTGCGAGCCCGGTGCACGCG	1620
Db	1549	GAGGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTGTGCGAGCCCGGTGCACGCG	1608
Qy	1621	GTGTACTACGACCCCGCAGAGGACCTGGTGGCCGAGATCCAGAGCAGGGCCGACGACCG	1680
Db	1609	GTGTACTACGACCCCGCAGAGGACCTGGTGGCCGAGATCCAGAGCAGGGCCGACGACCG	1668
Qy	1681	TGGACCTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG	1740
Db	1669	TGGACCTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG	1728
Qy	1741	ATGCGCACCCGCCACACCAACGACGTAAGCAGTGCACGAGCCGAGGCCGTGCAGAAATCGCC	1800
Db	1729	ATGCGCACCCGCCACACCAACGACGTAAGCAGTGCACGAGCCGAGGCCGTGCAGAAATCGCC	1788
Qy	1801	ATGAGAGCATCTGTGATCTGGGCGAGACCCCGCAAGTTCCGCTGCCATCTCAGAGGAG	1860
Db	1789	ATGAGAGCATCTGTGATCTGGGCGAGACCCCGCAAGTTCCGCTGCCATCTCAGAGGAG	1848
Qy	1861	ACCTGGGAGACCTGGTGGACCGACTACTGTGCAGGCCACTTGGATCCCGAGTGGAGTTC	1920
Db	1849	ACCTGGGAGACCTGGTGGACCGACTACTGTGCAGGCCACTTGGATCCCGAGTGGAGTTC	1908
Qy	1921	GTGAACACCCCGCTCTGGTGAAGCTGTGTTACGCTGGAGAGGAGGCCCATCATCGCC	1980
Db	1909	GTGAACACCCCGCTCTGGTGAAGCTGTGTTACGCTGGAGAGGAGGCCCATCATCGCC	1968
Qy	1981	GCCGAGACTTCTACGTGGAGGGCGCGCCGCAACCGCGAGACCAAGATCGGCAAGGCCGCG	2040
Db	1969	GCCGAGACTTCTACGTGGAGGGCGCGCGCCGCAACCGCGAGACCAAGATCGGCAAGGCCGCG	2028
Qy	2041	TACGTGACCGACGGGGCCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAG	2100
Db	2029	TACGTGACCGACGGGGCCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAG	2088
Qy	2101	ACCAGCTGCAGGCCCATCTAGCTGGCCCTGCAGGACAGCGGCGAGGCTGAACATCGT	2160



Db 2089 ACCGAGCTGCAGGCAATCCAGCTGGCCCTGCAGACACGCGCAGGTTGAAACATCGTG 2148  
 Qy 2161 ACCGACAGCCAGTAGTACGCCCTGGGATCATCCAGGCCCCAGCCCGACAAAGAGCGAGCGAG 2220  
 Db 2149 ACCGACAGCCAGTAGTACGCCCTGGGATCATCCAGGCCCCAGCCCGACAAAGAGCGAGCGAG 2208  
 Qy 2221 CTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAAAGGTGTACCTGAGCTGGGTG 2280  
 Db 2209 CTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAAAGGTGTACCTGAGCTGGGTG 2268  
 Qy 2281 CCCGCCACAGGCGCATCGCGGCAACGAGCAGATCGACAGCTGCTGAGCAAGGGGCATC 2340  
 Db 2269 CCCGCCACAGGCGCATCGCGGCAACGAGCAGATCGACAGCTGCTGAGCAAGGGGCATC 2328  
 Qy 2341 CGCAAGGTGCTGTTCTTCTGGAGCGCATCGATGGCGGCATCGTGATCTACCACTACAGGAC 2400  
 Db 2329 CGCAAGGTGCTGTTCTTCTGGAGCGCATCGATGGCGGCATCGTGATCTACCACTACAGGAC 2388  
 Qy 2401 GACCTGTACGTGGCAGCGCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2460  
 Db 2389 GACCTGTACGTGGCAGCGCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2448  
 Qy 2461 GGTGAATTC 2469  
 Db 2449 GGTGAATTC 2457

RESULT 11

ACA03546  
 ID ACA03546 standard; DNA; 2445 BP.  
 XX ACA03546;  
 AC ACA03546;  
 DT 22-MAY-2003 (first entry)  
 XX Synthetic DNA encoding immunogenic HIV peptide #29.  
 DE Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX Synthetic.  
 OS  
 XX WO2003004657-A1.  
 PN 16-JAN-2003.  
 XX 05-JUL-2002; 2002WO-US021421.  
 PF 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349733P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX (CHIR ) CHIRON CORP.  
 PA Zur Megede J, Barnett SW, Lian Y;  
 XX WPI; 2003-221602/21.  
 XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX Example 1; Fig 34; 262pp; English.  
 PS The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV

CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide  
 XX  
 Qy Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;  
 XX  
 Query Match 97.3%; Score 2401.8; DB 8; Length 2445;  
 Best Local Similarity 99.2%; Pred. No. 1.6e-289;  
 Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;  
 Qy 7 GCCACCATGGCCGAGGCGCATGAGCAGGCCACACAGCGCCACATCTCTGATCAGCGCAGC 66  
 Db 1 GCCACCATGGCCGAGGCGCATGAGCAGGCCACACAGCGCCACATCTCTGATGCGGCGCAGC 60  
 Qy 67 AACTTCAAGGGCCCAAGCGGATCATCAAGTGTCTTCACTGCGGCAAGAGGGGCCACATC 126  
 Db 61 AACTTCAAGGGCCCAAGCGGATCATCAAGTGTCTTCACTGCGGCAAGAGGGGCCACATC 120  
 Qy 127 GCCCGCAACTGCCCGCGCCCGCCGCAAGAGGGCTCTGGAAGTGGCGCAAGAGGGGCCAC 186  
 Db 121 GCCCGCAACTGCCCGCGCCCGCCGCAAGAGGGCTCTGGAAGTGGCGCAAGAGGGGCCAC 180  
 Qy 187 CAGATGAAGGACTGCACCCGAGCGCGCAGCCAACTTCTTCCCGAGGAGCTGGCTTTCCTCC 246  
 Db 181 CAGATGAAGGACTGCACCCGAGCGCGCAGCCAACTTCTTCCCGAGGAGCTGGCTTTCCTCC 240  
 Qy 247 CAGGGCAAGGCCCGCGAGTTTCCCGAGCAGCAGAACCCGCGCAACAGCCCCCACCAGCGCGC 306  
 Db 241 CAGGGCAAGGCCCGCGAGTTTCCCGAGCAGCAGAACCCGCGCAACAGCCCCCACCAGCGCGC 300  
 Qy 307 GAGCTCAGGTGGCGGGGAGCAACCCCGCAGCGAGGCGCGCGCGCGAGCGCGCGAGCGGACCC 366  
 Db 301 GAGCTCAGGTGGCGGGGAGCAACCCCGCAGCGAGGCGCGCGCGCGCGAGCGCGCGAGCGGACCC 360  
 Qy 367 CTGAATCTTCCCGCAGATCACCTGTGTGCGAGCGCGCCCTGTGTGAGCATCAAGTGTGGCGGC 426  
 Db 361 CTGAATCTTCCCGCAGATCACCTGTGTGCGAGCGCGCCCTGTGTGAGCATCAAGTGTGGCGGC 420  
 Qy 427 CAGATCAAGGAGGCCCTGTGTGAGCACCGCGCGCGCGACACCGTGTGTGAGAGAGATGAGC 486  
 Db 421 CAGATCAAGGAGGCCCTGTGTGAGCACCGCGCGCGCGACACCGTGTGTGAGAGAGATGAGC 480  
 Qy 487 CTGCCCGGCAAGTGGAAAGCCCAAGATCATCGGCGGCATCGGCGGCTTTCATCAAGGTGGCGC 546  
 Db 481 CTGCCCGGCAAGTGGAAAGCCCAAGATCATCGGCGGCATCGGCGGCTTTCATCAAGGTGGCGC 540  
 Qy 547 CAGTACGACAGATCTCTGTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCGTGTGTGATC 606  
 Db 541 CAGTACGACAGATCTCTGTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCGTGTGTGATC 600  
 Qy 607 GGCCCCACCCCCCGGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGTGACCCCTG 666  
 Db 601 GGCCCCACCCCCCGGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGTGACCCCTG 660  
 Qy 667 AACTTCCCGATCAGCCCCCATCGAGACCGTGGCCCGTGAAGCTGAAGCGCGGATGAGCGGC 726  
 Db 661 AACTTCCCGATCAGCCCCCATCGAGACCGTGGCCCGTGAAGCTGAAGCGCGGATGAGCGGC 720  
 Qy 727 CCAGAGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCGCTTACCGCGCATCTGCG 786  
 Db 721 CCAGAGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCGCTTACCGCGCATCTGCG 780  
 Qy 787 GAGGAGATGAGAGGAGGGGCAAGATCAACAGATCGGCGCGCGCGAGAACCCCTCAACACACC 846  
 Db 781 GAGGAGATGAGAGGAGGGGCAAGATCAACAGATCGGCGCGCGCGAGAACCCCTCAACACACC 840  
 Qy 847 CCGGTGTTCGCCATCAAGAGAGAGGAGCAGCACCAAGTGGCGCGCAAGCTGTGTGAGCTTCGC 906

Db 841 CCCGTGTTGCCATCAAGAAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900  
QY 907 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGGAGTGGAGTGGCATCCCGACCCGCGC 966  
Db 901 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGGAGTGGCATCCCGACCCGCGC 960  
QY 967 GGCCTGAAGAAGAAAGAGCGGTGACCGTGTCTGAGAGTGGCGCAGCGCTACTTTCAGCGTG 1026  
Db 961 GGCCTGAAGAAGAAAGAGCGGTGACCGTGTCTGAGAGTGGCGCAGCGCTACTTTCAGCGTG 1020  
QY 1027 CCCTTGGAACGAGACTTCCCAAGTACACCGCTTTTCAACATCCCGAGCATCAACAACGAG 1086  
Db 1021 CCCTTGAGCAGGACTTCCCAAGTACACCGCTTTTCAACATCCCGAGCATCAACAACGAG 1080  
QY 1087 ACCCCGGGATCCGCTACAGTACAAAGTCTGCTGCGCCCGAGGGCTGGAGGGCAGCCCGAC 1146  
Db 1081 ACCCCGGGATCCGCTACAGTACAAAGTCTGCTGCGCCCGAGGGCTGGAGGGCAGCCCGAC 1140  
QY 1147 ATCTTCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200  
QY 1207 GTGATCTACAGTACATGTCGACGCTGTGTACCTGGGACGACCTGTGGAGATCGGCGACGAC 1266  
Db 1201 GTGATCTACCA-----GGCCCCCTGTACGTGGGACGACCTGTGGAGATCGGCGACGAC 1254  
QY 1267 CGGCGCAAGATCGAGGAGCTGGCGAAGCACCTGCTGGGCTGGGGCTTCAACACCCCGAC 1326  
Db 1255 CGGCGCAAGATCGAGGAGCTGGCGAAGCACCTGCTGGGCTGGGGCTTCAACACCCCGAC 1314  
QY 1327 AAGAAGCACCAGAAAGAGGAGCCCTTCTGCTGTGTGGGTACGAGCTGACACCCCGCAAG 1386  
Db 1315 AAGAAGCACCAGAAAGAGGAGCCCTTCTGCTGTGTGGGTACGAGCTGACACCCCGCAAG 1368  
QY 1387 TGACCGGTGACGCCATCGAGCTGCCGAGAAGAGAGTGGACCGTGAACGACATCCAG 1446  
Db 1369 TGACCGGTGACGCCATCGAGCTGCCGAGAAGAGAGTGGACCGTGAACGACATCCAG 1428  
QY 1447 AAGCTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGCAG 1506  
Db 1429 AAGCTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGCAG 1488  
QY 1507 CTGTGAAGCTGTCTGCGCGCGCCCAAGGCCCTGACCGACATCTGTCGCCCTGACCGAGGAG 1566  
Db 1489 CTGTGAAGCTGTCTGCGCGCGCCCAAGGCCCTGACCGACATCTGTCGCCCTGACCGAGGAG 1548  
QY 1567 GCCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTGTGCGAGCCCGTGCACCGGCGTGTAC 1626  
Db 1549 GCCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTGTGCGAGCCCGTGCACCGGCGTGTAC 1608  
QY 1627 TAGACCCCGAGAGGACCTGTGGCGGAGATCCAGAGCAGGGCCGACGACGATGGACC 1686  
Db 1609 TAGACCCCGAGAGGACCTGTGGCGGAGATCCAGAGCAGGGCCGACGACGATGGACC 1668  
QY 1687 TACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGC 1746  
Db 1669 TACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGC 1728  
QY 1747 ACCGCCCCACCAACGACGTGAAGAGCTGACCGAGGCCGTGAGAAAGATCGCCATGGAG 1806  
Db 1729 ACCGCCCCACCAACGACGTGAAGAGCTGACCGAGGCCGTGAGAAAGATCGCCATGGAG 1788  
QY 1807 AGATCGTGTATCTGGGCGACAGACCCCAAGTTCCGCTGCGCATCCAGAGGAGACTGG 1866  
Db 1789 AGATCGTGTATCTGGGCGACAGACCCCAAGTTCCGCTGCGCATCCAGAGGAGACTGG 1848  
QY 1867 GAGACCTGTGGACCGGACTACTGTGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAAC 1926  
Db 1849 GAGACCTGTGGACCGGACTACTGTGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAAC 1908  
QY 1927 ACCCCCCCTGTGTGAAGCTGTGTATCCAGCTGGAGAAAGAGCCCATCATCTGGCGCGCAG 1986  
Db 1909 ACCCCCCCTGTGTGAAGCTGTGTATCCAGCTGGAGAAAGAGCCCATCATCTGGCGCGCAG 1968

QY 1987 ACCTTCTAGTGAACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTG 2046  
Db 1969 ACCTTCTAGTGAACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTG 2028  
QY 2047 ACCGACCGGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAAAGACGAG 2106  
Db 2029 ACCGACCGGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAAAGACGAG 2088  
QY 2107 CTGACGGCCATCCAGTGGCCCTGACGAGCAGCGGCGAGGCTGAACATCGTGACCGAC 2166  
Db 2089 CTGACGGCCATCCAGTGGCCCTGACGAGCAGCGGCGAGGCTGAACATCGTGACCGAC 2148  
QY 2167 AGCAGTACGCTGGGCGATCATCCAGGCGCCAGCCGACAAAGAGCGAGCGAGCTGGTG 2226  
Db 2149 AGCAGTACGCTGGGCGATCATCCAGGCGCCAGCCGACAAAGAGCGAGCGAGCTGGTG 2208  
QY 2227 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGCCCGCC 2286  
Db 2209 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGCCCGCC 2268  
QY 2287 CACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2346  
Db 2269 CACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2328  
QY 2347 GTGCTGTTCTGAGCGCATCGATGGCGCATCGTGTATCTACAGTACATGGAGCAGCTG 2406  
Db 2329 GTGCTGTTCTGAGCGCATCGATGGCGCATCGTGTATCTACAGTACATGGAGCAGCTG 2388  
QY 2407 TAGCTGGGCGAGCGCGCCCTAGTAGTCAATAAAGCTTCCCGGGGCTAGCACCGGT 2463  
Db 2389 TAGCTGGGCGAGCGCGCCCTAGTAGTCAATAAAGCTTCCCGGGGCTAGCACCGGT 2445

## RESULT 12

ADCL13264  
ID ADCL13264 standard; DNA; 2445 BP.

XX AC ADCL13264;

XX AC ADCL13264; (first entry)

DT 18-DEC-2003 (first entry)

XX DNA of HIV construct p2Pol-opt-YMWM\_C SEQ ID NO 43.

DE expression cassette; HIV Gag; Env; Int; Nef; p15NaseH; Pol; Tat; Prot;

XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX Human immunodeficiency virus.

XX WO2003004620-A2.

PD 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.

PR 31-AUG-2001; 2001US-0316860P.

PR 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

PA (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2003-221593/21.

XX New expression cassette comprising a polynucleotide sequence encoding a

XX polypeptide including an HIV Gag, Env, Int, Nef, p15NaseH, Pol, Tat,

XX Prot, or Rev polypeptide, useful for immunization, or generating

XX packaging cell lines.

XX Disclosure; Fig 40; 301pp; English.

The invention relates to a novel expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p13KaseH, Pol, Tat, Prot, or Rev polypeptide. The novel expression cassette can be used to treat HIV type C by gene therapy or used in the development of a vaccine. The gene delivery vector is administered intramuscularly, intracocally, intranasally, subcutaneously, intradermally, transdermally, intravaginally, intrarectally, orally or intravenously. The expression cassette is useful for immunisation, generating packaging cell lines and producing HIV polypeptides. This polynucleotide sequence represents the DNA of an HIV type C related sequence of the invention.

Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

**Query Match** 97.3%; Score 2401.8; DB 10; Length 2445;

Best Local Similarity 99.2%; Pred. No. 1.6e-289;

Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

Qy	7	GCACCATGGCGAGGCCATGAGCCAGGCGCACCGGCCCAACATCTGATGACAGCGCAGC	66
Db	1	GCCACCATGGCGAGGCCATGAGCCAGGCGCACCGGCCCAACATCTGATGACAGCGCAGC	60
Qy	67	AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTTCAACTGCGGCAAGAGAGGGCCACATC	126
Db	61	AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTTCAACTGCGGCAAGAGAGGGCCACATC	120
Qy	127	GCCCGCAACTCGCGCGCCCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGAGGGCCAC	186
Db	121	GCCCGCAACTCGCGCGCCCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGAGGGCCAC	180
Qy	187	CAGATGAAGGACTGCACGAGCGGCCAGGCAACTTTCTTCGCGAGGACCTGCGCTTCGCC	246
Db	181	CAGATGAAGGACTGCACGAGCGGCCAGGCAACTTTCTTCGCGAGGACCTGCGCTTCGCC	240
Qy	247	CAGGGCAAGGCCCGCGAGTTTCCCAAGCAGCAGAGAAACCGCGCCAAACAGCCCCCACCAGCCGC	306
Db	241	CAGGGCAAGGCCCGCGAGTTTCCCAAGCAGCAGAGAAACCGCGCCAAACAGCCCCCACCAGCCGC	300
Qy	307	GAGCTGACAGTGGCGGGCGCAAAACCCCGCAGCAGAGCGCGCGCGCGCGCAGCGCCAGGGCACC	366
Db	301	GAGCTGACAGTGGCGGGCGCAAAACCCCGCGCAGCAGAGCGCGCGCGCGCGCAGGGCACC	360
Qy	367	CTGAATCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGATCAAGGTGGCGCGC	426
Db	361	CTGAATCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGATCAAGGTGGCGCGC	420
Qy	427	CAGATCAAGGAGGGCCTCTCTGGAACAACCGCGCGCGCAGCACCCGTGTGGAGGAGATGAGC	486
Db	421	CAGATCAAGGAGGGCCTCTCTGGAACAACCGCGCGCGCAGCACCCGTGTGGAGGAGATGAGC	480
Qy	487	CTGCCCGCAAGTGGAGAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCAAGGTGGCGC	546
Db	481	CTGCCCGCAAGTGGAGAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCAAGGTGGCGC	540
Qy	547	CAGTACGACCAAGTCCTCATCGAGATCTGCGCAAGAAGGCCATCGCACCGTGTCTGATC	606
Db	541	CAGTACGACCAAGTCCTCATCGAGATCTGCGCAAGAAGGCCATCGCACCGTGTCTGATC	600
Qy	607	GGCCCCACCCCGTGAAACATCATCGCGCGCAACATGCTGAACCCAGCTGGGCTGCACCCCTG	666
Db	601	GGCCCCACCCCGTGAAACATCATCGCGCGCAACATGCTGAACCCAGCTGGGCTGCACCCCTG	660
Qy	667	AACTTTCCCATCAGCCCCCATCGAGNACCGTGCCTGTGAAGCTGMAAGCCCGGCGATGACCGC	726
Db	661	AACTTTCCCATCAGCCCCCATCGAGNACCGTGCCTGTGAAGCTGMAAGCCCGGCGATGACCGC	720
Qy	727	CCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC	786
Db	721	CCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGC	780
Qy	787	GAGGATGGAGAAAGGGGCAAGATCAACAGATCGGCCCGCGAGAAACCCCTTCAACACC	846
Db	781	GAGGATGGAGAAAGGGGCAAGATCAACAGATCGGCCCGCGAGAAACCCCTTCAACACC	840





PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
PS Disclosure; Fig 7; 301pp; English.  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramuscularly, intravenously,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC type C related sequence of the invention.  
XX  
SQ Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;  
Query Match 96.98; Score 2393.2; DB 10; Length 3930;  
Best Local Similarity 99.28; Pred. No. 1.7e-288;  
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;  
QY 14 TGGCCGAGGCGCATGAGCCAGCCACCAAGCGCCCAACATCTCTGATGCGAGCGCAACTTCA 73  
DB 1487 TCGCCGAGGCGCATGAGCCAGCCACCAAGCGCCCAACATCTCTGATGCGAGCGCAACTTCA 1546  
QY 74 AGGCCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGGCCCATATCGCCGCA 133  
DB 1547 AGGCCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGGCCCATATCGCCGCA 1606  
QY 134 ACTGCGCGCGCCCGCAAGAGGGCTGTGGAAGTGCAGAGGGGCGCACAGATGA 193  
DB 1607 ACTGCGCGCGCCCGCAAGAGGGCTGTGGAAGTGCAGAGGGGCGCACAGATGA 1666  
QY 194 AGGACTGACCGAGCGCGAGGCGCAACTCTTCTCCGCGAGGACCTTGGCCCTCCCGCAGGGCA 253  
DB 1667 AGGACTGACCGAGCGCGAGGCGCAACTCTTCTCCGCGAGGACCTTGGCCCTCCCGCAGGGCA 1726  
QY 254 AGGCGCGCGAGTTCGCGAGCGAGAGAACCGCGCGCAACCGCCGCAAGCGCGGAGTGC 313  
DB 1727 AGGCGCGCGAGTTCGCGAGCGAGAGAACCGCGCGCAACCGCCGCAAGCGCGGAGTGC 1786  
QY 314 AGGTGCGCGCGCAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
DB 1787 AGGTGCGCGCGCAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
QY 374 TCCCCCAGATCACCTGTGCGAGCGCCCGCTGTGTGAGCATCAAGGTGGCGCGCGCAGATCA 433  
DB 1847 TCCCCCAGATCACCTGTGCGAGCGCCCGCTGTGTGAGCATCAAGGTGGCGCGCGCAGATCA 1906  
QY 434 AGGAGGCCCTGTGGACACCGCGCGCGCAACACCGTGTGTGAGGAGATGAGCTGCGCGG 493  
DB 1907 AGGAGGCCCTGTGGACTCGCGCGCGCGCAACCGTGTGTGAGGAGATGAGCTGCGCGG 1966  
QY 494 GCAAGTGGAGGCCCAAGATCATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCGCGCATCG 553  
DB 1967 GCAAGTGGAGGCCCAAGATCATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCGCGCATCG 2026  
QY 554 ACCAGATCTCTGATCGAGATCTGGGGCAAGAGGCCATCGGCAACCGTGTGTGATGGCGCCCA 613  
DB 2027 ACCAGATCTCTGATCGAGATCTGGGGCAAGAGGCCATCGGCAACCGTGTGTGATGGCGCCCA 2086  
QY 614 CCCCCGTGAACATCATCGCGCGCGCAACATGTGTGACCCAGCTGGGGTGCACCTGAACTTCC 673  
DB 2087 CCCCCGTGAACATCATCGCGCGCGCAACATGTGTGACCCAGCTGGGGTGCACCTGAACTTCC 2146  
QY 674 CCATCAGCCCCCATCGAGACCGTGTGCGGTGAGCTGGAAGCCCGCGCATGAGCGGGCCCCAAGG 733  
DB 2147 CCATCAGCCCCCATCGAGACCGTGTGCGGTGAGCTGGAAGCCCGCGCATGAGCGGGCCCCAAGG 2206  
QY 734 TGAAGCAGGTGGCCCTTGACCGGAGAGAGATCAAGGCGCTTGACCGCCATCTGCGAGGAGA 793

DB 2207 TGAAGCAGTGGCCCTTGACCGGAGGAAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGA 2266  
QY 794 TGGAGAGGAGGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTCAACACACCCCGGTGT 853  
DB 2267 TGGAGAGGAGGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTCAACACACCCCGGTGT 2326  
QY 854 TCGCCCATCAAGAAAGAGGAGCAGCACCAAGTGGCGCAAGCTGTGTGGACTTCCCGGAGCTGA 913  
DB 2327 TCGCCCATCAAGAAAGAGGAGCAGCACCAAGTGGCGCAAGCTGTGTGGACTTCCCGGAGCTGA 2386  
QY 914 ACAAGCGCACCCAGGACTTCTTGGAGGATGAGCTGGGCAATCCCGCCACCCCGCGGCTCTGA 973  
DB 2387 ACAAGCGCACCCAGGACTTCTTGGAGGATGAGCTGGGCAATCCCGCCACCCCGCGGCTCTGA 2446  
QY 974 AGAAGAGAGAGCGGTGACCGTGTGAGCTGGGCGAGCGCTACTTTCAGGCTGCCCTGG 1033  
DB 2447 AGAAGAGAGAGCGGTGACCGTGTGAGCTGGGCGAGCGCTACTTTCAGGCTGCCCTGG 2506  
QY 1034 ACCAGGACTTTCGCAAGTACACCGCCTTTCACCATCCCGCAGCATCAACAAACGAGACCCCG 1093  
DB 2507 ACCAGGACTTTCGCAAGTACACCGCCTTTCACCATCCCGCAGCATCAACAAACGAGACCCCG 2566  
QY 1094 GCATCCGCTACCAAGTACCAAGTGTGCTGCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153  
DB 2567 GCATCCGCTACCAAGTACCAAGTGTGCTGCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626  
QY 1154 AGAGCAGCATGACCAAGATCTTGGAGCCCTTTCGCGCGCGCAACCCCGAGATCGTGATCT 1213  
DB 2627 AGAGCAGCATGACCAAGATCTTGGAGCCCTTTCGCGCGCGCAACCCCGAGATCGTGATCT 2686  
QY 1214 ACCAGTACATGAGACCACTGTGCTGGGCGAGCAGCATCTGGAGATCGGCGAGCACCAGCGCCA 1273  
DB 2687 ACCA-----GGCCCCCTGTGCTGGGCGAGCAGCATCTGGAGATCGGCGAGCACCAGCGCCA 2740  
QY 1274 AGATCAGAGAGTGCAGCAAGCACTGTGCTGTGGGCTTTCACACCCCGCAACAAGAGC 1333  
DB 2741 AGATCAGAGAGTGCAGCAAGCACTGTGCTGTGGGCTTTCACACCCCGCAACAAGAGC 2800  
QY 1334 ACCAGAGAGAGCCCTTCTGTGGATGGGCTTTCAGAGCTGCAACCCCGCAAGTGGACCG 1393  
DB 2801 ACCAGAGAGAGCCCTTCTGTGGATGGGCTTTCAGAGCTGCAACCCCGCAAGTGGACCG 2854  
QY 1394 TGCAGGCCATTCAGAGCTGCCCGAAGAGAGAGCTGGACCGGTGAACGAGCATCCAGAAGCTGG 1453  
DB 2855 TGCAGGCCATTCAGAGCTGCCCGAAGAGAGAGCTGGACCGGTGAACGAGCATCCAGAAGCTGG 2914  
QY 1454 TGGGCAAGCTGAACCTGGGCGAGCAGATCTAACCAGGATCAAGGTGCGGCGAGCTGTGCA 1513  
DB 2915 TGGGCAAGCTGAACCTGGGCGAGCAGATCTAACCAGGATCAAGGTGCGGCGAGCTGTGCA 2974  
QY 1514 AGCTGTGCGCGCGCGCCAGAGCCCTGACCGACATCTGTGCGCCCTGACCGAGAGGCGCGAGC 1573  
DB 2975 AGCTGTGCGCGCGCGCCAGAGCCCTGACCGACATCTGTGCGCCCTGACCGAGAGGCGCGAGC 3034  
QY 1574 TGCAGCTGGCGCGAAGAACCGCGAGATCTTGCAGGAGCCGCTGACCGGCTGTACTACGACC 1633  
DB 3035 TGCAGCTGGCGCGAAGAACCGCGAGATCTTGCAGGAGCCGCTGACCGGCTGTACTACGACC 3094  
QY 1634 CAGCAAGGAGCCTGTGTGGCGAGATCTCAGAGAGAGGCGCAACGAGTGAACCTTACAGAG 1693  
DB 3095 CAGCAAGGAGCCTGTGTGGCGAGATCTCAGAGAGAGGCGCAACGAGTGAACCTTACAGAG 3154  
QY 1694 TCTTACAGGAGCCCTTTCAGAACTTGAAGCCCGGCAAGTACGCAAGATCGCGACCGCC 1753  
DB 3155 TCTTACAGGAGCCCTTTCAGAACTTGAAGCCCGGCAAGTACGCAAGATCGCGACCGCC 3214  
QY 1754 ACACCAACGAGCTGAAGAGCAGCTGACCGAGGCGCTGAGAGAGATCGCCATGAGAGCATCG 1813  
DB 3215 ACACCAACGAGCTGAAGAGCAGCTGACCGAGGCGCTGAGAGAGATCGCCATGAGAGCATCG 3274  
QY 1814 TGAATCTGGGCGCAAGACCCCGCAAGTTCGCTGCGCATTCAGAGAGGAGACCTTGGAGACCT 1873



Db 3275 TGATCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGACCTGGAGACCT 3334  
 Qy 1874 GGTGACCGACTACTGGCAGCCCACTGGATCCCGAGTGGGAGTTGCTGAACACCCCCC 1933  
 Db 3335 GGTGACCGACTACTGGCAGCCCACTGGATCCCGAGTGGGAGTTGCTGAACACCCCCC 3394  
 Qy 1934 CCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993  
 Db 3395 CCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454  
 Qy 1994 ACCTGGAGCGCGCCGCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTGACCGACC 2053  
 Db 3455 ACCTGGAGCGCGCGCCGCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTGACCGACC 3514  
 Qy 2054 GGGCGCGCGCAAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 2113  
 Db 3515 GGGCGCGCGCAAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 3574  
 Qy 2114 CCATCCAGCTGGCCCTGCAGACAGCGGCGAGCGAGGCTGAACATCGTGACGACGACG 2173  
 Db 3575 CCATCCAGCTGGCCCTGCAGACAGCGGCGAGCGAGGCTGAACATCGTGACGACGACG 3634  
 Qy 2174 ACGCCCTGGGCTCATCCAGGCGCCAGCCGCAAGAGCGAGCGAGCTGGTGAACAGA 2233  
 Db 3635 ACGCCCTGGGCTCATCCAGGCGCCAGCCGCAAGAGCGAGCGAGCTGGTGAACAGA 3694  
 Qy 2234 TCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTCTGAGCTGGGTGCCGCCCAAGG 2293  
 Db 3695 TCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTCTGAGCTGGGTGCCGCCCAAGG 3754  
 Qy 2294 GCATCGGCGGCAACGAGCAGATCGACAAAGCTGTGTGAGCAAGGGCATCGGCAAGGTGCTGT 2353  
 Db 3755 GCATCGGCGGCAACGAGCAGATCGACAAAGCTGTGTGAGCAAGGGCATCGGCAAGGTGCTGT 3814  
 Qy 2354 TCCTGGAGCGCATCGATGGCGGCATCGTGTACTACCAAGTACATGAGACCTGTACCTGG 2413  
 Db 3815 TCCTGGAGCGCATCGATGGCGGCATCGTGTACTACCAAGTACATGAGACCTGTACCTGG 3874  
 Qy 2414 GCAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463  
 Db 3875 GCAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 15  
 ID ADC13232  
 X AC ADC13232;  
 X DE 18-DEC-2003 (first entry)  
 X DNA of HIV construct GagCompIPolmutIna\_c SEQ ID NO 11.  
 X expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 X Rev; HIV type C; Gene therapy; vaccine; immunisation; HIV; ds.  
 X Human immunodeficiency virus.  
 X WO2003004620-A2.  
 X 16-JAN-2003.  
 X 05-JUL-2002; 2002WO-US021420.  
 X 05-JUL-2001; 2001US-0303192P.  
 X 31-AUG-2001; 2001US-0316860P.  
 X 16-JAN-2002; 2002US-0349871P.  
 X (CHIR ) CHIRON CORP.  
 X (UYST-) UNIV STELLENBOSCH.  
 X Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 X

DR WPI; 2003-221593/21.  
 XX New expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide, useful for immunization, or generating packaging cell lines.  
 PT Disclosure; Fig 8; 301pp; English.  
 PS The invention relates to a novel expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel expression cassette can be used to treat HIV type C by gene therapy or used in the development of a vaccine. The gene delivery vector is administered intramuscularly, intramuscularly, intravenously, subcutaneously, orally or intravenously. The expression cassette is useful for immunisation, generating packaging cell lines and producing HIV polypeptides. This polynucleotide sequence represents the DNA of an HIV Type C related sequence of the invention.  
 XX Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;  
 SQ  
 Query Match 96.9%; Score 2393.2; DB 10; Length 3930;  
 Best Local Similarity 99.2%; Pred. No. 1.7e-288;  
 Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;  
 Qy 14 TGGCCGAGGCGCATGAGCGAGCCACAGCGCCCAACATCTCTGATGCGAGCAACTTCA 73  
 Db 1487 TCGCCGAGGCGCATGAGCGAGCCACAGCGCCCAACATCTCTGATGCGAGCAACTTCA 1546  
 Qy 74 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGCCACATGCCCGCA 133  
 Db 1547 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGCCACATGCCCGCA 1606  
 Qy 134 ACTGCGCGCGCCCGCAAGAGGGCTCTTGGAGTGGCGGAGGGCCACAGATGA 193  
 Db 1607 ACTGCGCGCGCCCGCAAGAGGGCTCTTGGAGTGGCGGAGGGCCACAGATGA 1666  
 Qy 194 AGGACTGACCGAGCGCGAGCGCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGAGGCA 253  
 Db 1667 AGGACTGACCGAGCGCGAGCGCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGAGGCA 1726  
 Qy 254 AGGCCCGCGAGTTCCCGAGCGAGCAACCGCGCAACAGCCCCACAGCGCGAGCTGC 313  
 Db 1727 AGGCCCGCGAGTTCCCGAGCGAGCAACCGCGCAACAGCCCCACAGCGCGAGCTGC 1786  
 Qy 314 AGGTGCGCGGAGCAACCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 Db 1787 AGGTGCGCGGAGCAACCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
 Qy 374 TCCCGCAGATCACCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGTGGCGCGCGAGATCA 433  
 Db 1847 TCCCGCAGATCACCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGTGGCGCGCGAGATCA 1906  
 Qy 434 AGGAGGCCCTGTGTGACACCGCGCGCGCGAGCAACCGCGCGAGAGATGAGCTGCCCG 493  
 Db 1907 AGGAGGCCCTGTGTGACACCGCGCGCGCGAGCAACCGCGCGAGAGATGAGCTGCCCG 1966  
 Qy 494 GCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGCGCTTCAATCAAGTGGCGCGAGTACG 553  
 Db 1967 GCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGCGCTTCAATCAAGTGGCGCGAGTACG 2026  
 Qy 554 ACCAGATCTGTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCGCTGTGTGATCGGCGCA 613  
 Db 2027 ACCAGATCTGTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCGCTGTGTGATCGGCGCA 2086  
 Qy 614 CCCCCTGGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCC 673  
 Db 2087 CCCCCTGGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCC 2146  
 Qy 674 CCATCAGCCCCATCGAGACCGTGCCTGAGCTGAAGCTGAAGCCCGCGCATGAGCGCGCCCAAGG 733

```

Db 2147 CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGAGACGGCCCCAAGG 2206
QY 734 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGCCCTGACCGCCATCTGCGAGAGA 793
Db 2207 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGCCCTGACCGCCATCTGCGAGAGA 2266
QY 794 TGAAGAGGAGGCGAAGATCACCAAGATCGGCCCGAGAAACCCCTACAAACCCCGGTGT 853
Db 2267 TGAAGAGGAGGCGAAGATCACCAAGATCGGCCCGAGAAACCCCTACAAACCCCGGTGT 2326
QY 854 TGCCTCATCAAGAAAGAGAGACACCAAGTGGGCGAAGCTGTGTGAGATCTCCGCGAGCTGA 913
Db 2327 TGCCTCATCAAGAAAGAGAGACACCAAGTGGGCGAAGCTGTGTGAGATCTCCGCGAGCTGA 2386
QY 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGCATCCCCACCGCCGCGGCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGCATCCCCACCGCCGCGGCTGA 2446
QY 974 AGAAGAGAGAGCGTGAACCGTGTCTGGACGTGGGCGACGCTACTTTCAGCGTGCCTCTGG 1033
Db 2447 AGAAGAGAGAGCGTGAACCGTGTCTGGACGTGGGCGACGCTACTTTCAGCGTGCCTCTGG 2506
QY 1034 ACAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAACGAGACCCCG 1093
Db 2507 ACAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAACGAGACCCCG 2566
QY 1094 GCATCCGCTACAGTACAAAGTCTGCCAGGCTGGAAGGCGAGCCACCGATCTTCC 1153
Db 2567 GCATCCGCTACAGTACAAAGTCTGCCAGGCTGGAAGGCGAGCCACCGATCTTCC 2626
QY 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686
QY 1214 ACCAGTACATGGAACGACTGTGAGTGGGCGAGACGACCTGAGATCGGCGAGCACCGGCCA 1273
Db 2687 ACCA-----GGCCCCCTGTAGCTGGGCGAGCAGCCTGAGATCGGCGCAGCACCGGCCA 2740
QY 1274 AGATCGAGGCTGCCAGACACTCTGCGCTGGGGCTTACACCCCGCAAGAGC 1333
Db 2741 AGATCGAGGCTGCCAGACACTCTGCGCTGGGGCTTACACCCCGCAAGAGC 2800
QY 1334 ACCAGAGAGACCCCTCTCTGTGATGGGCTACGAGCTGCAACCCCGCAAGTGGACCG 1393
Db 2801 ACCAGAGAGACCCCTCTCTGTGATGGGCTACGAGCTGCAACCCCGCAAGTGGACCG 2854
QY 1394 TGCAGCCCATCGAGTGCCTGCGAGAGAGCTGGAACCGTGAACGATCGAAGCTGG 1453
Db 2855 TGCAGCCCATCGAGTGCCTGCGAGAGAGCTGGAACCGTGAACGATCGAAGCTGG 2914
QY 1454 TGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCA 1513
Db 2915 TGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCA 2974
QY 1514 AGCTGTGCGCGCGCCAGGCGCTGACCGACATCTGTGCCCTGACCGAGGCGCGAGC 1573
Db 2975 AGCTGTGCGCGCGCCAGGCGCTGACCGACATCTGTGCCCTGACCGAGGCGCGAGC 3034
QY 1574 TGAAGTGGCGGAGAACCGCGAGATCTGTGCGAGCGCGCTGCAAGCGGTGTATACGACC 1633
Db 3035 TGAAGTGGCGGAGAACCGCGAGATCTGTGCGAGCGCGCTGCAAGCGGTGTATACGACC 3094
QY 1634 CCAGCAGGACCTGGTGGCGAGATCCAGAGCAGGCGCCAGCAGCTGAGCTTACCAGA 1693
Db 3095 CCAGCAGGACCTGGTGGCGAGATCCAGAGCAGGCGCCAGCAGCTGAGCTTACCAGA 3154
QY 1694 TCTACAGAGACCTTCAAGAACTGAAAGCCGCGCAAGTACGCGCAAGATCGCAACCGCCC 1753
Db 3155 TCTACAGAGACCTTCAAGAACTGAAAGCCGCGCAAGTACGCGCAAGATCGCAACCGCCC 3214
QY 1754 ACACCAACGAGCTGAGAGCAGCTGACCGAGCGCTGCGAAGAGATCGCCATGAGAGCATCG 1813
Db 3215 ACACCAACGAGCTGAGAGCAGCTGACCGAGCGCTGCGAAGAGATCGCCATGAGAGCATCG 3274

```

```

QY 1814 TGATCTGGGGCAAGACCCCAAGTTTCCGCTTCCATCCAGAAAGGAGACCTGGGAGACCT 1873
Db 3275 TGATCTGGGGCAAGACCCCAAGTTTCCGCTTCCATCCAGAAAGGAGACCTGGGAGACCT 3334
QY 1874 GGTGGAACGACTACTGCGAGGCGACCTGATCCCGAGTGGGAGTTTCTGTGAACACCCCC 1933
Db 3335 GGTGGAACGACTACTGCGAGGCGACCTGATCCCGAGTGGGAGTTTCTGTGAACACCCCC 3394
QY 1934 CCTGTGTGAAGCTGTGGTACCAGTGGAGAAAGAGCCCATCATCGGCGCGGAGACCTTCT 1993
Db 3395 CCTGTGTGAAGCTGTGGTACCAGTGGAGAAAGAGCCCATCATCGGCGCGGAGACCTTCT 3454
QY 1994 AGTGTGACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACC 2053
Db 3455 AGTGTGACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACC 3514
QY 2054 GGGGCGGCGAGAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 2113
Db 3515 GGGGCGGCGAGAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 3574
QY 2114 CCATCCAGCTGGCCCTGACAGGACAGCGGAGGTGAACATCTGTGACCGAGCCAGT 2173
Db 3575 CCATCCAGCTGGCCCTGACAGGACAGCGGAGGTGAACATCTGTGACCGAGCCAGT 3634
QY 2174 AGCCCTGGGCGATCATCCAGGCGCCAGCCGACAAAGAGCGAGAGCTGGTGAACCGAGA 2233
Db 3635 AGCCCTGGGCGATCATCCAGGCGCCAGCCGACAAAGAGCGAGAGCTGGTGAACCGAGA 3694
QY 2234 TCATCAGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCCCGCCACAGG 2293
Db 3695 TCATCAGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCCCGCCACAGG 3754
QY 2294 GCATCGGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTGT 2353
Db 3755 GCATCGGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTGT 3814
QY 2354 TCTGTGACGGCGATCGATGGCGGCATCTGTGATCTTACCAAGTACATGAGACGCTGTACGTGG 2413
Db 3815 TCTGTGACGGCGATCGATGGCGGCATCTGTGATCTTACCAAGTACATGAGACGCTGTACGTGG 3874
QY 2414 GCMGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db 3875 GCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

```

Search completed: June 1, 2005, 11:33:18  
 Job time : 1280.42 secs





Db 181 GGCCACCAAGTGAAGACTGACACGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240  
Qy 241 TTCCCCACAGGCAAGGCCCGGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCCACC 300  
Db 241 TTCCCCACAGGCAAGGCCCGGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCCACC 300  
Qy 301 AGCGCGAGCTGACAGGTGCGCGGCGACAAACCCCGCAGCAGAGCGCGCGCGAGCGCCAG 360  
Db 301 AGCGCGAGCTGACAGGTGCGCGGCGACAAACCCCGCAGCAGAGCGCGCGCGAGCGCCAG 360  
Qy 361 GGCAACCTGAACTTCCCCAGATCAACCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTG 420  
Db 361 GGCAACCTGAACTTCCCCAGATCAACCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTG 420  
Qy 421 GGCGGCCAGATCAAGAGGCCCTGTGGACACCGGCGCCGACACACCTGTGTGGAGGAG 480  
Db 421 GGCGGCCAGATCAAGAGGCCCTGTGGACACCGGCGCCGACACACCTGTGTGGAGGAG 480  
Qy 481 ATGAGCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAG 540  
Db 481 ATGAGCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAG 540  
Qy 541 GTGCGCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG 600  
Db 541 GTGCGCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG 600  
Qy 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Qy 661 ACCCTGAACTTCCCATCAGCCCATCGACACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720  
Db 661 ACCCTGAACTTCCCATCAGCCCATCGACACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720  
Qy 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
Qy 781 ATCTGGAGAGATGGAAGAGAGGGCAAGATCAACAGATCGGCCCCGAGAACCCCTTAC 840  
Db 781 ATCTGGAGAGATGGAAGAGAGGGCAAGATCAACAGATCGGCCCCGAGAACCCCTTAC 840  
Qy 841 AACACCCCGCTGTTCGCACTCAAGAAAGACAGACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGCTGTTCGCACTCAAGAAAGACAGACCAAGTGGCGCAAGCTGGTGGAC 900  
Qy 901 TTCGGGAGCTGAACAAAGCGCACCCAGACTTCTGGGAGGTGACGTGGGCATCCCCCAC 960  
Db 901 TTCGGGAGCTGAACAAAGCGCACCCAGACTTCTGGGAGGTGACGTGGGCATCCCCCAC 960  
Qy 961 CCGCGGCTGGAAGAGAGAGGTGACCGTGTGACCTGGGCGACGCTACTTC 1020  
Db 961 CCGCGGCTGGAAGAGAGAGGTGACCGTGTGACCTGGGCGACGCTACTTC 1020  
Qy 1021 AGCGTGCCCTGACGAGGACTTCCGCAAGTACACCGCTTCAACCATCCCGAGCATCAAC 1080  
Db 1021 AGCGTGCCCTGACGAGGACTTCCGCAAGTACACCGCTTCAACCATCCCGAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGGCATCCGCTACAGTAAACCGTGTGCTGCCCGAGGGCTGGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGGCATCCGCTACAGTAAACCGTGTGCTGCCCGAGGGCTGGAAGGGCAGC 1140  
Qy 1141 CCAGCATCTTCAGAGCAGATGACCAAGATCTGGAGCCCTTCGCGCCCGCAACCCC 1200  
Db 1141 CCAGCATCTTCAGAGCAGATGACCAAGATCTGGAGCCCTTCGCGCCCGCAACCCC 1200  
Qy 1201 GAGATCGTGTACTACAGTACATGAGACCTGTGCTGGGCGAGCAGCTGGAGATCGGC 1260  
Db 1201 GAGATCGTGTACTACAGTACATGAGACCTGTGCTGGGCGAGCAGCTGGAGATCGGC 1260  
Qy 1261 CAGCACCGGCCAAGATCGAGGAGCTGCGCAAGCAGCTGTGGCTGGGGCTTCAACACC 1320  
Db 1261 CAGCACCGGCCAAGATCGAGGAGCTGCGCAAGCAGCTGTGGCTGGGGCTTCAACACC 1320

Qy 1321 CCCGACAAAGAACACACAGAAAGGAGCCCCCTTCTGTGTGATGGCTACGAGCTGCACCCC 1380  
Db 1321 CCCGACAAAGAACACACAGAAAGGAGCCCCCTTCTGTGTGATGGCTACGAGCTGCACCCC 1380  
Qy 1381 GACAAGTGGAACCGTGCAGCGCCATCGAGCTGCCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440  
Db 1381 GACAAGTGGAACCGTGCAGCGCCATCGAGCTGCCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440  
Qy 1441 ATCCAGAACTGTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500  
Db 1441 ATCCAGAACTGTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500  
Qy 1501 CGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGCGCTGACCGACATCTGTGCCCTGACC 1560  
Db 1501 CGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGCGCTGACCGACATCTGTGCCCTGACC 1560  
Qy 1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAAACCGCGAGATCTGCGCGAGCCCTGTGCACGCG 1620  
Db 1561 GAGGAGGCCGAGCTGGAGCTGGCCGAGAAACCGCGAGATCTGCGCGAGCCCTGTGCACGCG 1620  
Qy 1621 GTGTACTACGACCCGAGAAAGGACCTGTGTGGCCGAGATCCAGAAAGCAGGGCCAGCCAG 1680  
Db 1621 GTGTACTACGACCCGAGAAAGGACCTGTGTGGCCGAGATCCAGAAAGCAGGGCCAGCCAG 1680  
Qy 1681 TGGACCTTACCAGATCTACCGAGGAGCCCTTCAAGAAACCTGAAGACCGGCAAGTACGCCAAG 1740  
Db 1681 TGGACCTTACCAGATCTACCGAGGAGCCCTTCAAGAAACCTGAAGACCGGCAAGTACGCCAAG 1740  
Qy 1741 ATGCGCACCGCCCAACCAACGACGTAAGACGCTGACCGAGGCCGTGCAAGATCGCC 1800  
Db 1741 ATGCGCACCGCCCAACCAACGACGTAAGACGCTGACCGAGGCCGTGCAAGATCGCC 1800  
Qy 1801 ATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCCCTGCCATCCAGAAAGGAG 1860  
Db 1801 ATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCCCTGCCATCCAGAAAGGAG 1860  
Qy 1861 ACCTGGAGACCTGTGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTTC 1920  
Db 1861 ACCTGGAGACCTGTGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTTC 1920  
Qy 1921 GTGAAACACCCCCCTGTGTGAAAGCTGTGTGTACAGCTGGAGAAGGAGCCCATCATCGGC 1980  
Db 1921 GTGAAACACCCCCCTGTGTGAAAGCTGTGTGTACAGCTGGAGAAGGAGCCCATCATCGGC 1980  
Qy 1981 GCGGAGACCTTCTACGTGACCGCGCGGCCAAACCGCGAGACCAAGATCGGCAAGGCCGCG 2040  
Db 1981 GCGGAGACCTTCTACGTGAGCGCGCGGCCAAACCGCGAGACCAAGATCGGCAAGGCCGCG 2040  
Qy 2041 TAGCTGACCGACCGGGCGCGGCAAGATCGTGTGAGCTGTGACCGAGACCAACCAAGGAG 2100  
Db 2041 TAGCTGACCGACCGGGCGCGGCAAGATCGTGTGAGCTGTGACCGAGACCAACCAAGGAG 2100  
Qy 2101 ACCGAGCTGACGCCATTCAGCTGGCCCTGACGAGACGCGGACGAGGTGAAACATCGTG 2160  
Db 2101 ACCGAGCTGACGCCATTCAGCTGGCCCTGACGAGACGCGGACGAGGTGAAACATCGTG 2160  
Qy 2161 ACCGACAGCCAGTACGCGCTGGGCAATCATCGGCCCGCCGACAGAGGAGGAGCGAG 2220  
Db 2161 ACCGACAGCCAGTACGCGCTGGGCAATCATCGGCCCGCCGACAGAGGAGGAGCGAG 2220  
Qy 2221 CTGCTGAAACAGATCATCGAGCAGCTGTATCAAGAAAGGAGAGGTGTACTCTGAGCTGGGTG 2280  
Db 2221 CTGCTGAAACAGATCATCGAGCAGCTGTATCAAGAAAGGAGAGGTGTACTCTGAGCTGGGTG 2280  
Qy 2281 CCCGCCCAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGGATC 2340  
Db 2281 CCCGCCCAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGGATC 2340  
Qy 2341 CGCAAGGTGCTGTCTCTGACCGGCATCGATGGCGGCATCGTGTATCTACCAAGTACATGGAC 2400  
Db 2341 CGCAAGGTGCTGTCTCTGACCGGCATCGATGGCGGCATCGTGTATCTACCAAGTACATGGAC 2400

Qy 2401 GACCTGTACGTGGGAGGCGGCGCTAGGATCGATTAAAGCTTCCGGGGCTAGCAC 2460  
Db 2401 GACCTGTACGTGGGAGGCGGCGCTAGGATCGATTAAAGCTTCCGGGGCTAGCAC 2460  
Qy 2461 GGTGAATTC 2469  
Db 2461 GGTGAATTC 2469

RESULT 2  
AX455915 2463 bp DNA linear PAT 06-JUL-2002  
LOCUS  
DEFINITION Sequence 31 from Patent WO0204493.  
ACCESSION AX455915  
VERSION AX455915.1 GI:21714900  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
other sequences; artificial sequences.

REFERENCE 1  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES  
Location/Qualifiers  
source 1..2463  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YM"

ORIGIN  
Query Match 98.9%; Score 2442.2; DB 6; Length 2463;  
Best Local Similarity 99.6%; Pred. No. 1.6e-246;  
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
Qy 1 GTCAGCCCAACATGGCCGAGGCATGAGCAGGCCACAGCGCCAAATCCTCTGATGAG 60  
Db 1 GTCAGCCCAACATGGCCGAGGCATGAGCAGGCCACAGCGCCAAATCCTCTGATGAG 60  
Qy 61 CGGAGCACTTCAGGCGCCCAAGCGCATCATCAAGTGTCTCACTGGCGGAGGAGGC 120  
Db 61 CGGAGCACTTCAGGCGCCCAAGCGCATCATCAAGTGTCTCACTGGCGGAGGAGGC 120  
Qy 121 CACATCGCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Qy 181 GGCCACAGATGAAGGACTGACCGAGCGCCAGGCCAACTTCTTCGCGAGGAGCCTGGCC 240  
Db 181 GGCCACAGATGAAGGACTGACCGAGCGCCAGGCCAACTTCTTCGCGAGGAGCCTGGCC 240  
Qy 241 TTTCCCGGAGGCGCGCGAGTTCCTCAGCGAGCAGACCGCGCCCAAGCGCCCGCCACC 300  
Db 241 TTTCCCGGAGGCGCGCGAGTTCCTCAGCGAGCAGACCGCGCCCAAGCGCCCGCCACC 300  
Qy 301 AGCGCGAGCTGCAAGTGGCGGCGCAACCCCGCAGCGAGGCGCGCGCGCGCGAG 360  
Db 301 AGCGCGAGCTGCAAGTGGCGGCGCAACCCCGCAGCGAGGCGCGCGCGCGCGAG 360  
Qy 361 GGCAACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGGTG 420  
Db 361 GGCAACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGGTG 420  
Qy 421 GGCGGCGAGATCAAGGAGGCGCTCTGACACCGCGCGCGCGAGCACCGTGTCTGGAGGAG 480  
Db 421 GGCGGCGAGATCAAGGAGGCGCTCTGACACCGCGCGCGCGAGCACCGTGTCTGGAGGAG 480  
Qy 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCGCATCGCGGCTTCATCAAG 540  
Db 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCGCATCGCGGCTTCATCAAG 540

Qy 541 GTGCGCCAGTACGACGAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGACCGGTG 600  
Db 541 GTGCGCCAGTACGACGAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGACCGGTG 600  
Qy 601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Qy 661 ACCCTGAATTCCTCCCATCAGCCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGATG 720  
Db 661 ACCCTGAATTCCTCCCATCAGCCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGATG 720  
Qy 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCC 780  
Db 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCC 780  
Qy 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACAGATCGGCGCCCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACAGATCGGCGCCCGAGAACCCCTAC 840  
Qy 841 AACACCCCGTGTTCGCCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGGTGAC 900  
Db 841 AACACCCCGTGTTCGCCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGGTGAC 900  
Qy 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGCC 960  
Db 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGCC 960  
Qy 961 CCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGAGCGTGGGCGAGCGCTTACTTC 1020  
Db 961 CCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGAGCGTGGGCGAGCGCTTACTTC 1020  
Qy 1021 AGCGTGCCTTGGACGAGGAGCTTCCGCAAGTACACCGCTTTCACCATCTCCCGAGCATCAAC 1080  
Db 1021 AGCGTGCCTTGGACGAGGAGCTTCCGCAAGTACACCGCTTTCACCATCTCCCGAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGCGCTACCGTACAGTACAACTGCTGCCCGCAGGCTGGAAGGCGAGC 1140  
Db 1081 AACGAGACCCCGCGCTACCGTACAGTACAACTGCTGCCCGCAGGCTGGAAGGCGAGC 1140  
Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGGAGCGCTTCCGCGCGCGCAACCC 1200  
Db 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGGAGCGCTTCCGCGCGCGCAACCC 1200  
Qy 1201 GAGATCGTGTACTACAGTATATGGACGACCTGTACGTGGGCGAGCGACTGGAGATCGGC 1260  
Db 1201 GAGATCGTGTACTACCA- - - - -GGCGCGCTGTACGTGGGCGAGCGACTGGAGATCGGC 1260  
Qy 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAGCAGCACTGCTGCGCTGGGCTTACCGACC 1320  
Db 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAGCAGCACTGCTGCGCTGGGCTTACCGACC 1320  
Qy 1321 CCGGCAAGAGCAGCAAGAGGAGCGCCCTTCTGCTGGATGGGCTACGAGCTGCACCGCC 1380  
Db 1321 CCGGCAAGAGCAGCAAGAGGAGCGCCCTTCTGCTGGATGGGCTACGAGCTGCACCGCC 1380  
Qy 1381 GACAAGTGGACCGCTGACGCGCATCGAGCTGCGCGAGAGGAGCTGAGCGCTGGAACGAC 1440  
Db 1381 GACAAGTGGACCGCTGACGCGCATCGAGCTGCGCGAGAGGAGCTGAGCGCTGGAACGAC 1440  
Qy 1441 ATCCAGAGCTGTGGGCAAGTGAATCGGCGCAGCAGATCTACCGCGGATCAAGGTG 1500  
Db 1441 ATCCAGAGCTGTGGGCAAGTGAATCGGCGCAGCAGATCTACCGCGGATCAAGGTG 1500  
Qy 1501 CGCGAGCTGTGAGCTGTGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
Db 1501 CGCGAGCTGTGAGCTGTGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
Qy 1561 GAGGAGCGCGAGCTGTGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620  
Db 1561 GAGGAGCGCGAGCTGTGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620  
Qy 1621 GTGTACTTACGACCCCAAGGAGCTGTGGTGGCGGAGATCCAGAAAGAGGCGCGCGAG 1680  
Db 1621 GTGTACTTACGACCCCAAGGAGCTGTGGTGGCGGAGATCCAGAAAGAGGCGCGCGAG 1680

Db 1615 GTGTACTACGACCCAGCAGGACCTGGTGGCGAGATCCAGAAGCAGGCGCCACGACCAG 1674  
Qy 1681 TGGACCTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAG 1740  
Db 1675 TGGACCTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAG 1734  
Qy 1741 ATGGCACCGCCACACCAAGCAGCTGAAGCAGCTGACCCAGCGCGTGCAGAGATCGCC 1800  
Db 1735 ATGGCACCGCCACACCAAGCAGCTGAAGCAGCTGACCCAGCGCGTGCAGAGATCGCC 1794  
Qy 1801 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCGCTCCATCCAGAAGGAG 1860  
Db 1795 ATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCGCTCCATCCAGAAGGAG 1854  
Qy 1861 ACCTGGAGACCTGGTGGACCGACTACTGCGAGCCCACTTGGATCCCGAGTGGAGTTC 1920  
Db 1855 ACCTGGAGACCTGGTGGACCGACTACTGCGAGCCCACTTGGATCCCGAGTGGAGTTC 1914  
Qy 1921 GTCAACACCCCGCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGC 1980  
Db 1915 GTCAACACCCCGCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGC 1974  
Qy 1981 GCCGAGACCTTCACTGCTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGC 2040  
Db 1975 GCCGAGACCTTCACTGCTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGC 2034  
Qy 2041 TAGCTGACCAACCGGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAG 2100  
Db 2035 TAGCTGACCAACCGGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAG 2094  
Qy 2101 ACCGAGCTGACGCGCATCCAGCTGGCCCTGACGAGACCGGCGAGCGAGTGAACATCGTG 2160  
Db 2095 ACCGAGCTGACGCGCATCCAGCTGGCCCTGACGAGACCGGCGAGCGAGTGAACATCGTG 2154  
Qy 2161 ACCGAGACCGAGTACGCTGGGCGATCATCCAGCGCCAGCCCGACCAAGAGCGAGCGAG 2220  
Db 2155 ACCGAGACCGAGTACGCTGGGCGATCATCCAGCGCCAGCCCGACCAAGAGCGAGCGAG 2214  
Qy 2221 CTGGTGAAACAGATCATCGAGCGCTGATCAAGAGAGAGTGTACCTGAGCTGGGTG 2280  
Db 2215 CTGGTGAAACAGATCATCGAGCGCTGATCAAGAGAGAGTGTACCTGAGCTGGGTG 2274  
Qy 2281 CCCGCCACAAAGGCGCATCGGCGCAACAGCAGATCGACAAGCTGGTGACAAGGCGCATC 2340  
Db 2275 CCCGCCACAAAGGCGCATCGGCGCAACAGCAGATCGACAAGCTGGTGACAAGGCGCATC 2334  
Qy 2341 CGCAAGTGCTGTTCTTGGACGGCGATCGATGGCGGCATCGTGATCTACCAAGTACATGGAC 2400  
Db 2335 CGCAAGTGCTGTTCTTGGACGGCGATCGATGGCGGCATCGTGATCTACCAAGTACATGGAC 2394  
Qy 2401 GACTGTACCTGGGAGCGCGGCTTACGATCGATTAAGCTTCCCGGGCTAGCAC 2460  
Db 2395 GACTGTACCTGGGAGCGCGGCTTACGATCGATTAAGCTTCCCGGGCTAGCAC 2454  
Qy 2461 GGTGAATTC 2469  
Db 2455 GGTGAATTC 2463

RESULT 3  
AX455916  
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 32 from Patent W00204493.  
ACCESSION AX455916  
VERSION AX455916.1 GI:21714901  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
other sequences; artificial sequences.  
REFERENCE 1  
zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides.

polypeptides and uses thereof  
Patent: WO 0204493-A 32 17-JAN-2002;  
CHIRON CORPORATION (US) : University of Stellenbosch (ZA)  
FEATURES  
source Location/Qualifiers  
1. .2457  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YMW"  
ORIGIN  
Query Match 97.8%; Score 2415.4; DB 6; Length 2457;  
Best Local Similarity 99.3%; Pred. No. 1e-243;  
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;  
Qy 1 GTCAGCGCCACCATGCGCGAGGCCATGAGCCAGCGCCACCGAGCCACACATCTCTGATCGAG 60  
Db 1 GTCAGCGCCACCATGCGCGAGGCCATGAGCCAGCGCCACCGAGCCACACATCTCTGATCGAG 60  
Qy 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120  
Db 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120  
Qy 121 CACATCGCCGCAACTTCCCGCGCCCGCCCGCAAGAGGGCTGTGGAAAGTGTGGCAAGGAG 180  
Db 121 CACATCGCCGCAACTTCCCGCGCCCGCCCGCAAGAGGGCTGTGGAAAGTGTGGCAAGGAG 180  
Qy 181 GGCACAGATGAAGACTGACCGAGCGCGCAGCCCAACTTCTTCGCGAGGACCTGGCC 240  
Db 181 GGCACAGATGAAGACTGACCGAGCGCGCAGCCCAACTTCTTCGCGAGGACCTGGCC 240  
Qy 241 TTCCCGCAGGCAAGCGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCCCACC 300  
Db 241 TTCCCGCAGGCAAGCGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCCCACC 300  
Qy 301 AGCGCGAGCTGCAAGTGTGCGCGCGACAACCCCCGAGCGAGCGCGCGCGCGCGCGAG 360  
Db 301 AGCGCGAGCTGCAAGTGTGCGCGCGACAACCCCCGAGCGAGCGCGCGCGCGCGCGAG 360  
Qy 361 GGCACCGTGAACCTTCCCGCAGATCACCTGTGCGAGGCGCCCTGTGTGAGCATCAAGGTG 420  
Db 361 GGCACCGTGAACCTTCCCGCAGATCACCTGTGCGAGGCGCCCTGTGTGAGCATCAAGGTG 420  
Qy 421 GCGGCGCAGATCAAGGAGGCGCTGTGGACACCGCGCGCGACACACCGTGTGGAGGAG 480  
Db 421 GCGGCGCAGATCAAGGAGGCGCTGTGGACACCGCGCGCGACACACCGTGTGGAGGAG 480  
Qy 481 ATGAGCGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540  
Db 481 ATGAGCGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540  
Qy 541 GTGCGCGAGTACACCAAGATCTGTGATGAGATCTGCGGCAAGAGGCGCATCGGCGCGGTG 600  
Db 541 GTGCGCGAGTACACCAAGATCTGTGATGAGATCTGCGGCAAGAGGCGCATCGGCGCGGTG 600  
Qy 601 CTGATCGGCTGC 660  
Db 601 CTGATCGGCTGC 660  
Qy 661 ACCCTGAACCTTCCCGCATCGAGCCCGCATCGAGACCGTGCCTGGAAGCTGAAGCGCGCATG 720  
Db 661 ACCCTGAACCTTCCCGCATCGAGCCCGCATCGAGACCGTGCCTGGAAGCTGAAGCGCGCATG 720  
Qy 721 GACGCC 780  
Db 721 GACGCC 780  
Qy 781 ATCTGCGAGAGATGGAGAGGAGGCGCAAGATCAACAGATCGCGCGCGCGCGCGCGCGCTAC 840  
Db 781 ATCTGCGAGAGATGGAGAGGAGGCGCAAGATCAACAGATCGCGCGCGCGCGCGCGCGCTAC 840  
Qy 841 AACACCCCGCTGTTCGCCATCAAGAGAGGAGCAGCACCAAGTGTGGCGCAAGCTGTGGAC 900  
Db 841 AACACCCCGCTGTTCGCCATCAAGAGAGGAGCAGCACCAAGTGTGGCGCAAGCTGTGGAC 900

841 AACACCCCGCTGTTCCGCATCAAGAGAAAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCATCCCCAC 960  
901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCATCCCCAC 960  
961 CCAGCGGCTGAAGAGAAAGAGCGTGAAGCTGCTGGAGCTGGGCGAGCGCTACTTC 1020  
961 CCAGCGGCTGAAGAGAAAGAGCGTGAAGCTGCTGGAGCTGGGCGAGCGCTACTTC 1020  
1021 AGCGTGCCTTGGAGAGAGACTTCCGAAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
1021 AGCGTGCCTTGGAGAGAGACTTCCGAAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
1081 AACGAGACCCCGGATCCCGTACAGTACAACTGCTGCCCGAGGCTGGAAGGGCAGC 1140  
1081 AACGAGACCCCGGATCCCGTACAGTACAACTGCTGCCCGAGGCTGGAAGGGCAGC 1140  
1141 CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCCCGCAACCCC 1200  
1141 CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCCCGCAACCCC 1200  
1201 GAGATCGTGTACTACCAAGTACATGAGACCTGTAGTGGCGAGCGACTGGAGATCGGC 1260  
1201 GAGATCGTGTACTACCAAGTACATGAGACCTGTAGTGGCGAGCGACTGGAGATCGGC 1260  
1201 GAGATCGTGTACTACCAAGTACATGAGACCTGTAGTGGCGAGCGACTGGAGATCGGC 1260  
1261 CAGCACCGCGCAAGATGAGAGCTGCGCAAGCACTGCTGCGCTGGGGCTTCCACACC 1320  
1255 CAGCACCGCGCAAGATGAGAGCTGCGCAAGCACTGCTGCGCTGGGGCTTCCACACC 1314  
1321 CCGACAAAGACCAAGAGAGCGCCCTTCTGTGGATGGGCTAGAGCTGCACCCC 1380  
1315 CCGACAAAGACCAAGAGAGCGCCCTTCTGTGCCAT- - - - -CGAGCTGCACCCC 1368  
1381 GACAAGTGGACCGTGCAGCCCATGAGCTGCCGAGAGGAGAGCTGACCGTGAACGAC 1440  
1369 GACAAGTGGACCGTGCAGCCCATGAGCTGCCGAGAGGAGAGCTGACCGTGAACGAC 1428  
1441 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1500  
1429 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1488  
1501 CGCAGCTGTCAAGCTGTCTGGGCGCAGGCGCTGACGACATGTCCTGAC 1560  
1489 CGCAGCTGTCAAGCTGTCTGGGCGCAGGCGCTGACGACATGTCCTGAC 1548  
1561 GAGGAGCGGAGCTGGAGCTGGCGGAGAACCGGAGATCTGCGGAGCGCGTGCAGCGC 1620  
1549 GAGGAGCGGAGCTGGAGCTGGCGGAGAACCGGAGATCTGCGGAGCGCGTGCAGCGC 1608  
1621 GTGTACTACGACCCAGCAAGACTTGTGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1680  
1609 GTGTACTACGACCCAGCAAGACTTGTGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1668  
1681 TGGACCTACAGATCTACAGGAGCGCTTCAAGAACTGGAAGACCGGCAAGTACGCCAAG 1740  
1669 TGGACCTACAGATCTACAGGAGCGCTTCAAGAACTGGAAGACCGGCAAGTACGCCAAG 1728  
1741 ATGCGCACCGCCACACCAAGAGTGAAGAGCTGACCGAGCGCTGCAAGAGATCGCC 1800  
1729 ATGCGCACCGCCACACCAAGAGTGAAGAGCTGACCGAGCGCTGCAAGAGATCGCC 1788  
1801 ATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAG 1860  
1789 ATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAG 1848  
1861 ACTTGGAGAGCTGTGTGAACCACTATGCGAGCGCACCTGATGATCCCGAGTGGAGTTC 1920  
1849 ACTTGGAGAGCTGTGTGAACCACTATGCGAGCGCACCTGATGATCCCGAGTGGAGTTC 1908  
1921 GTGAACACCCCGCTGTTGAGCTGTGTTGAGCTGTGTTGAGAGAGGAGCCCATCATCGGC 1980  
1909 GTGAACACCCCGCTGTTGAGCTGTGTTGAGCTGTGTTGAGAGAGGAGCCCATCATCGGC 1968

QY 1981 GCCGAGACCTTCTTACTGTTGACGCGCGCCCAACCGCGAGACAAGATCGGCAAGCGCGC 2040  
Db 1969 GCCGAGACCTTCTTACTGTTGACGCGCGCCCAACCGCGAGACAAGATCGGCAAGCGCGC 2028  
QY 2041 TACGTGACCGACCGCGCGCGCAGAGATCTGTAGCGCTGTAGCGGACCAACCAACGAGAG 2100  
Db 2029 TACGTGACCGACCGCGCGCGCAGAGATCTGTAGCGCTGTAGCGGACCAACCAACGAGAG 2088  
QY 2101 ACCGAGCTGACGAGGCATCCAGCTGGCCCTGAGGACAGCGCGCAGCGAGTGAACATCGTG 2160  
Db 2089 ACCGAGCTGACGAGGCATCCAGCTGGCCCTGAGGACAGCGCGCAGCGAGTGAACATCGTG 2148  
QY 2161 ACCGAGCGCAGTACGCTTGGGCTATCATCCAGCGCCAGCGCCGACCAAGAGCGAGCGAG 2220  
Db 2149 ACCGAGCGCAGTACGCTTGGGCTATCATCCAGCGCCAGCGCCGACCAAGAGCGAGCGAG 2208  
QY 2221 CTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGGTG 2280  
Db 2209 CTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGGTG 2268  
QY 2281 CCGCGCCACAGGCGCATCGCGGCAACGAGCAGATCGACAAAGCTGGTGAAGAGGCGATC 2340  
Db 2269 CCGCGCCACAGGCGCATCGCGGCAACGAGCAGATCGACAAAGCTGGTGAAGAGGCGATC 2328  
QY 2341 CGCAAGGTGCTTCTTGGAGCGCATCGATCGCGCATCGTGTATCTACGATACATGGAC 2400  
Db 2329 CGCAAGGTGCTTCTTGGAGCGCATCGATCGCGCATCGTGTATCTACGATACATGGAC 2388  
QY 2401 GACCTGTACGTGGGCGCGCGCTAGGATCGATTAAGAGCTTCCCGGGCTAGCACC 2460  
Db 2389 GACCTGTACGTGGGCGCGCGCTAGGATCGATTAAGAGCTTCCCGGGCTAGCACC 2448  
QY 2461 GGTGAATTC 2469  
Db 2449 GGTGAATTC 2457

RESULT 4  
BD263706  
LOCUS  
DEFINITION BD263706 2312 bp DNA linear PAT 17-JUL-2003  
Improved expression of HIV polypeptides and production of  
virus-like particles.  
ACCESSION BD263706  
VERSION BD263706.1 GI:33073474  
KEYWORDS JP 2002533124-A/73.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 2312)  
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,  
Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: JP 2002533124-A 73 08-OCT-2002;  
CHIRON CORP  
COMMENT OS Artificial Sequence  
PN JP 2002533124-A/73  
PD 08-OCT-2002  
PF 30-DEC-1999 JP 2000591193  
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI  
SUSAN BARNETT,JAN ZUR MESEDE,INDRESH SRIVASTAVA,YING LIAN, PI  
KARIN HARTOG,  
PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC  
CI2N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC  
CI2N5/10,  
PC CI2N7/00,CI2P21/02,CI2N15/00,CI2N5/00,A61K37/02 CC  
Description of Artificial Sequence: FS(-).protmod.RTopt(+) FH Key  
Location/Qualifiers  
FT source 1..2312  
FT /organism='Artificial Sequence'.  
Location/Qualifiers  
source 1..2312

/organism="synthetic construct"		83.1%; Score 2052; DB 6; Length 2312;	
/mol_type="genomic DNA"		93.6%; Pred. No. 9.6e-206;	
/db_xref="taxon:32630"		0; Mismatches 135; Indels 12; Gaps 2;	
ORIGIN			
Query Match	Best Local Similarity	Matches 2165; Conservative	
170	GCGGCAAGAGGCGCCACCAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGC	229	
1	GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCGC	60	
230	AGGACTTGGCTTCCCGAGGCAAGGCGCGAGTTCGCCAGCGAGCAGAAACCGCGCCA	289	
61	AGGACTTGGCTTCTCGAGGCGAAGGCGCGAGTTCAGCAGCGAGCAGACCGCGCCA	120	
290	ACAGCCCCACACGCGCGAGCTCAGGTGCGCGGCG-----ACAACCCCGCGAGG	343	
121	ACAGCCCCACCGCGAGCTCAGGTGCGGCGCGAGAAACAACAGCCTGAGCGAGG	180	
344	CGGCGCGGAGCGCCAGGCGACCTG-----AACTTCCCGGAGATCACCTCTGGCAGC	397	
181	CGGCGCGGACCGCCAGGCGACCGTGAGCTTCAACTTCCCGGAGATCACCTCTGGCAGC	240	
398	GCCCCCTGTGAGCATCAAGGTGGCGGCGCAGATCAAGGAGGCGCTGCTGGACACCGGCG	457	
241	GCCCCCTGTGTACCATCAGGATCGGCGGCGCAGCTCAAGGAGGCGCTGCTGCACACCGGCG	300	
458	CGACGACACCGTGTGGAGGAGATGAGCTGCGCGGCAAGTGGAAAGCCCAAGATGATCG	517	
301	CGACGACACCGTGTGGAGGAGTGAACCTGCGCGGCAAGTGGAAAGCCCAAGATGATCG	360	
518	GCGGATCGGCGGCTTCATCAAGGTGGCGGCGCAGTACGACAGATCTTGATCGAGATCTGG	577	
361	GCGGATCGGCGGCTTCATCAAGGTGGCGGCGCAGTACGACAGATCTTGATCGAGATCTGG	420	
578	GCAAGAGGCGCATCGGCACCGTGTGATCGGCGGCCACCCCGGTGAACATCATCGGCGCA	637	
421	GCCAAAGGCGCATCGGCACCGTGTGTTGGCGGCCACCCCGGTGAACATCATCGGCGCA	480	
638	ACATGTGTACCCAGCTGGGCTGCACCTGAACTTCCCATATCAGCCCCATCGAGACCGTGC	697	
481	ACCTGTGTACCCAGATCGGCTGCACCTGAACTTCCCATATCAGCCCCATCGAGACCGTGC	540	
698	CGTGAAGCTGAAGCCCGGATGGAACCGGCCCAAGGTGAAGAGTGGCCCTTACCGAGG	757	
541	CGTGAAGCTGAAGCCCGGATGGAACCGGCCCAAGGTGAAGAGTGGCCCTTACCGAGG	600	
758	AGAAGATCAAGGCGCTTACCGCATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCA	817	
601	AGAAGATCAAGGCGCTTGTGGAGATCTGACCGAGATGGAGAGGAGGCGCAAGATCAGCA	660	
818	AGATCGGCGCGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAGAGGACAGCA	877	
661	AGATCGGCGCGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAGAGGACAGCA	720	
878	CAAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGG	937	
721	CAAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGG	780	
938	AGGTGCAAGCTGGGCGATCCCGCACCCCGCGGCTTGAAGAAGAGAGCGTGACCGTGC	997	
781	AGGTGCAAGCTGGGCGATCCCGCACCCCGCGGCTTGAAGAAGAGAGCGTGACCGTGC	840	
998	TGACGTGGGCGAGCGCTACTTACGCGTGCCTTCAGCGAGGACTTCCGCAAGTACACCG	1057	
841	TGACGTGGGCGAGCGCTACTTACGCGTGCCTTCAGCGAGGACTTCCGCAAGTACACCG	900	
1058	CTTTACCATCCCGAGGATCAACAAAGAGACCCCGCGCATCCCGTACCAAGTACAAAGTGC	1117	
901	CTTTACCATCCCGAGGATCAACAAAGAGACCCCGCGCATCCCGTACCAAGTACAAAGTGC	960	

Qy	1118	TGCCCCAGGGCTGGAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG	1117
Db	961	TGCCCCAGGGCTGGAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG	1020
Qy	1178	AGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCTACCACTATCATGGACGACCTGTACG	1237
Db	1021	AGCCCTTCCGCAAGCAGAAACCCCGACATCTGTATCTTACCAGTATCATGACGACCTGTACG	1080
Qy	1238	TGGGCAAGCACTTGAGATTCGGCCAGCAACCGCGCCAAAGATCGAGGAGCTGGCGAAGCAAC	1297
Db	1081	TGGGCAAGCACTTGAGATTCGGCCAGCAACCGCACCAAGATCGAGGAGCTGGCGCAGCAC	1140
Qy	1298	TGCTGCGCTGGGGCTTCCACCCCGCAAGAAGACCAAGAGAGAGCCCTTCTCTGT	1357
Db	1141	TGCTGCGCTGGGGCTTCCACCCCGCAAGAAGACCAAGAGAGAGCCCTTCTCTGT	1200
Qy	1358	GGATGGGCTACGAGCTGCAACCCCGCAAAAGTGGACCGTGACGCCATCGAGCTGCCGAGA	1417
Db	1201	GGATGGGCTACGAGCTGCAACCCCGCAAAAGTGGACCGTGACGCCATCATGCTGCCGAGA	1266
Qy	1418	AGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGGCAGCC	1477
Db	1261	AGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGGCAGCC	1320
Qy	1478	AGATCTAACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCC	1537
Db	1321	AGATCTAACCCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCCCAAGGCC	1380
Qy	1538	TGACCGACATCTGTGCCCTTGACCGAGGAGCGCAGCTGGAGCTGGCCGAGAAACCGCGAGA	1597
Db	1381	TGACCGAGGTGATTCCTCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAGAAACCGCGAGA	1440
Qy	1598	TCCTGCGCAGCCCTGTCACGCGCTGTACTACGCCCCAGCAAGGACCTGTGTGGCCGAGA	1657
Db	1441	TCCTGAAGAGGCCGTGTGACGAGGTGTACTACGCCCCAGCAAGGACCTGTGTGGCCGAGA	1500
Qy	1658	TCAGAAGCAGGGGCCACGACCAAGTGGACCTTACCAGATCTACGAGAGCCCTTCAAGAACC	1717
Db	1501	TCAGAAGCAGGGGCCAGGGCCAGTGGACCTTACCAGATCTACGAGAGCCCTTCAAGAACC	1560
Qy	1718	TGAAGACCGGCAAGTACGCCAAGATCGGCAACGCCCAACCAACGACGTGAAGCAGCTGA	1777
Db	1561	TGAAGACCGGCAAGTACGCCCGCATTCGCGCGCGCCACCAACGACGTGAAGCAGCTGA	1620
Qy	1778	CGAGGCCCTGCAGAAGATTCGCCATTCGAGAGCATCGTGTCTGGGGCAAGACCCCCCAAGT	1837
Db	1621	CGAGGCCCTGCAGAAGGTGAGCAACGAGAGCATCTGTATCTGGGGCAAGATCCCCCAAGT	1680
Qy	1838	TCGCGCTGCCATTCAGAAGAGGAGACTCTGGGAGACTGTGTGAGCCGACTACTTGGCAGGCA	1897
Db	1681	TCAGCTGCCATTCAGAAGAGGAGACTCTGGGAGGCCCTGGTGGATGGAGTACTTGGCAGGCA	1740
Qy	1898	CCTGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTGGTGAAGCTGTGTGTTACCAAG	1957
Db	1741	CCTGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTGGTGAAGCTGTGTGTTACCAAG	1800
Qy	1958	TGGAGAAGGAGGCCATCATTCGCGCGCGAGACTTCTACGTGAGCGCGCCGCCAACCCGCG	2017
Db	1801	TGGAGAAGGAGGCCATTCGTTGGCGCGGAGACTTCTACGTGAGCGCGCCGCCAACCCGCG	1866
Qy	2018	AGACCAAGATCGGCAAGCGCGCTAGTGAACGACCGGGGCGCGCAGAGAGTCTGTGAGCC	2077
Db	1861	AGACCAAGCTGGGCAAGCGCGCTAGTGAACGACCGGGGCGCGCAGAGAGTCTGTGAGCA	1920
Qy	2078	TGACCGAGACCAACCAAGAGACCGAGCTTCAGGCCCATCCAGCTGGCCCTTCGAGGACA	2137
Db	1921	TGCGCCGACCAACCAAGAGACCGAGCTTCAGGCCCATCCACCTGSCCTTCGAGGACA	1980
Qy	2138	GCGGCAAGGAGTGAACATCTGTGAACGACGACGATGACGCGCGCGCGCTTCGAGGCC	2197
Db	1981	GCGGCTGGAGGTGAACATCTGTGAACGACGACGATGACGCGCGCGCGCTTCGAGGCC	2040
Qy	2198	AGCTCCACAGAGCGGAGCGGAGCTTGTGTGAACCAACATCATCGAGCAGCTGATCAAGAAGG	2257

QY	1118	TGCCCCAGGGCTGGAAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGG	1177
DB	961	TGCCCCAGGGCTGGAAAGGGCAGCCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCTGG	1020
QY	1178	AGCCCTTTCGGCGCCGCAACCCCGAGATCGTGATCTTACAGTACATGGAGCAGCTGTACG	1237
DB	1021	AGCCCTTTCGGCAAGCAGAAACCCCGACATCGTGATCTTACCAGTACATGGAGCAGCTGTACG	1080
QY	1238	TGGGAGCAGCAGCTTGGAGATCGGCCAGCACCGGCCCAAGATCGAGGAGCTCGCGCAAGCAC	1297
DB	1081	TGGGAGCAGCAGCTTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTCGCGCAGCAC	1140
QY	1298	TGCTGGCTGGGGCTTTCACACCCCGCAACAAGAACCAAGAGAGCCCTTCTCTGT	1357
DB	1141	TGCTGGCTGGGGCTTTCACACCCCGCAACAAGAACCAAGAGAGCCCTTCTCTGT	1200
QY	1358	GGATGGCTACGAGCTGCAACCCCGCAACAAGTGGACCGTGGCCCAAGCTGAGCTGCGCGAGA	1417
DB	1201	GGATGGCTACGAGCTGCAACCCCGCAACAAGTGGACCGTGGCCCAAGCTGAGCTGCGCGAGA	1260
QY	1418	AGGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC	1477
DB	1261	AGGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC	1320
QY	1478	AGATCTACCCCGGCTTCAAGGTGGCGGCTGTGCAAGTGTGCGCGGCGCCCAAGGCC	1537
DB	1321	AGATCTACCCCGGCTTCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGGCAACCAAGGCC	1380
QY	1538	TGACCGACATCTGTCCTTGAACCGAGGCGCGAGCTGGAGCTGGGCGGAGAACCGCGAGA	1597
DB	1381	TGACCGAGGTGATTCCTTGTGACCGAGGCGCGAGCTGGAGCTGGGCGGAGAACCGCGAGA	1440
QY	1598	TCCTGCGCAGCCCGTGCAGCGGCTGTACTACACCCCGCAAGAGGACCTGGTGGCGCGAGA	1657
DB	1441	TCCTGAAGAGCCCGTGCACGAGGTGTACTACACCCCGCAAGAGGACCTGGTGGCGCGAGA	1500
QY	1658	TCAGAAAGCAGGCGCCACGACCTTACAGATCTTACAGAGGAGCCCTTCAAGAAC	1717
DB	1501	TCAGAAAGCAGGCGCCACGAGCTTACAGATCTTACAGAGGAGCCCTTCAAGAAC	1560
QY	1718	TGAAGACCGGCAAGTACGCGACAGTGGCCACCGCCACACCAACGAGGTGAAGCAGCTGA	1777
DB	1561	TGAAGACCGGCAAGTACGCGCGCATCGCGCGCCACCAACACGAGCTGAAGCAGCTGA	1620
QY	1778	CCGAGCCGTGCAGAAAGATCGCCATCGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGT	1837
DB	1621	CCGAGCCGTGCAGAAAGTGAACACCGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGT	1680
QY	1838	TCCGCTTCCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGGACTACTGGCAGGCCA	1897
DB	1681	TCAAGCTGCCCATCCAGAAAGGAGACCTGGGAGGCTGGTGGATGGAGTACTGGCAGGCCA	1740
QY	1898	CTTGGATCCCGAGTGGGAGTTCTGAAACACCCCGCCCTGGTGAAGCTGTGTACCGAGC	1957
DB	1741	CTTGGATCCCGAGTGGGAGTTCTGAAACACCCCGCCCTGGTGAAGCTGTGTACCGAGC	1800
QY	1958	TGAGAAAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGGCGCCCAACCGCG	2017
DB	1801	TGAGAAAGGAGCCCATCTGTGGGCGCGAGACCTTCTACGTGGAGCGGCGCCCAACCGCG	1860
QY	2018	AGACCAAGTTCGGCAAGGCGGCTACGTGACCGAGCCGCGCGAGAGGAGGTTGTGAGCA	2077
DB	1861	AGACCAAGTTCGGCAAGGCGGCTACGTGACCGAGCCGCGCGAGAGGAGGTTGTGAGCA	1920
QY	2078	TGACCGAGACCAACCAAGAGACCGAGCTGACGGCCATCCAGCTGGCCCTCGAGACA	2137
DB	1921	TCGCCGACACCAACCAAGAGACCGAGCTGACGGCCATCCACCTGGCCCTCGAGACA	1980
QY	2138	GCGGAGCAGAGGTGAACATCGTGACCGAGCAGTACGCCCTGGGGCATCATCAGGCCCC	2197
DB	1981	GCGGCTTGGAGGTGAACATCGTGACCGAGCAGTACGCCCTGGGGCATCATCAGGCCCC	2040
QY	2198	AGCCCGACAAAGAGCGAGCGAGCTGGTGAAACAGATCTATCGAGAGCTGATCAAGAGG	2257



Db 2041 AGCCCGACAGAGCGAGCGAGCTGGTGAGCCAGATCATGAGAGCTGATCAAGAAGG 2100  
Qy 2258 AGAAGGTGTACTGAGCTGGGTGCGCGCCCAACAAGGGCATCGCGCGCAACAGAGCAGATCG 2317  
Db 2101 AGAAGGTGTACTGCGCTGGGTGCGCGCCCAACAAGGGCATCGCGCGCAACAGAGCAGTGG 2160  
Qy 2318 ACAAGCTGGTGAGCAAGGGCATCGCAAGGTGTCTTCCTGGAGCGGATCGATCGCGGCA 2377  
Db 2161 ACAAGCTGGTGAGCGCGCGCATCGCAAGGTGTCTTCCTGAAACGGATCGATCGCGGCA 2220  
Qy 2378 TCGTGATCTACAGTACATGAGCAGCCTGTACGTGGGCGAGCGCGGCGCTAGGATCGATT 2437  
Db 2221 TCGTGATCTACAGTACATGAGCAGCCTGTACGTGGGCGAGCGCGGCGCTAGGATCGATT 2280  
Qy 2438 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2469  
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 5  
LOCUS CQ870576 2312 bp DNA linear PAT 13-SEP-2004  
DEFINITION Sequence 84 from Patent EP1433851.  
ACCESSION CQ870576  
VERSION CQ870576.1 GI:52000092  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
other sequences; artificial sequences.

REFERENCE  
1 Barnett,S., Zumegede,J., Srivastava,I., Lian,Y., Hartog,K.,  
AUTHORS Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: EP 1433951-A 84 30-JUN-2004;  
CHIRON CORPORATION (US)  
FEATURES  
source  
1. .2312  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/notes="Description of Artificial Sequence:  
FS(-).protmod.Rtopt(+)"

ORIGIN

Query Match 83.1%; Score 2052; DB 6; Length 2312;  
Best Local Similarity 93.6%; Pred. No. 9.6e-206;  
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

Qy 170 GCGGCAAGGAGGGCCACCAGATGAAGACTGCGACCGAGCGCGCCAACTTCTTCGCG 229  
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGCAGGCTAATTCTTCGCG 60  
Qy 230 AGAAGCTGGCTTCCCCCAGGCGAAGCGCGCGAGTTCCCGAGGAGCAGAAACCGCGCCA 289  
Db 61 AGAAGCTGGCTTCTCGAGGCAAGCGCGCGAGTTTCAAGCAGGAGCAGACCGCGCCA 120  
Qy 290 ACAGCCCCACAGCGCGAGCTGAGGTGGCGCGCG-----ACAACCCCGGAGCGAG 343  
Db 121 ACAGCCCCACCGCGCGAGCTGAGGTGGCGCGCGAGGAGAGAAACAAGCTGAGCGAG 180  
Qy 344 CCGGCGCGAGCGCGAGCGGCAACCTG-----AACTTCCCCCAGATCACCTGTGGCAGC 397  
Db 181 CCGGCGCGAGCGCGAGCGGCAACCTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC 240  
Qy 398 GCGGCTGGTGAGATCAAGGTGGCGCGCAGATCAAGAGGAGCGCTGCTGAGACACCGCG 457  
Db 241 GCGGCTGGTGAGATCAAGGTGGCGCGCAGCTCAAGAGGAGCGCTGCTGAGACACCGCG 300  
Qy 458 CCGAGCACCGGTGCTGGAGAGATGAGCTGCGCGCAAGTGGAGCGGCAAGATGATCG 517  
Db 301 CCGAGCACCGGTGCTGGAGAGATGAGCTGCGCGCAAGTGGAGCGGCAAGATGATCG 360

Qy 518 GCGCATCGCGGCTTCAATCAAGGTGCGCCAGTAGACCAAGATCCTGATCAGATCTGCG 577  
Db 361 GCGGATCGCGGCTTCAATCAAGGTGCGCGAGTAGACCAAGATCCTGATCAGATCTGCG 420  
Qy 578 GCAAGAGGCGCATCGGACCGTGTGATCGGCGCCCAACCGCTGAACATCATCGCGCGCA 637  
Db 421 GCCAAGGCGCATCGGACCGTGTGATCGGCGCCCAACCGCTGAACATCATCGCGCGCA 480  
Qy 638 ACATGCTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGC 697  
Db 481 ACCTGCTGACCCAGATCGGCTGCACCCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGC 540  
Qy 698 CCGTGAAGCTGAAGCCCGGCGATGGAACGCGCCCAAGGTGAAGCATGTGCGCCCTGACCGAGG 757  
Db 541 CCGTGAAGCTGAAGCCCGGCGATGGAACGCGCCCAAGGTGAAGCATGTGCGCCCTGACCGAGG 600  
Qy 758 AGAAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGGAAGAGGCGCAAGATCACCA 817  
Db 601 AGAAGATCAAGGCGCTGCTGGAGATGCGACCGAGATGGAAGAGGCGCAAGATCACCA 660  
Qy 818 AGATCGGCGCGAGAACCCCTTACAACACCCCGCTGTTCCGCATCAAGAAGAGGACAGCA 877  
Db 661 AGATCGGCGCGAGAACCCCTTACAACACCCCGCTGTTCCGCATCAAGAAGAGGACAGCA 720  
Qy 878 CCAAGTGGCGAAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 937  
Db 721 CCAAGTGGCGAAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 780  
Qy 938 AGTGCAGCTGGGCGATCCCGCCCGCGCGCTGGAAGAGAGGCGTGCAGCTGC 997  
Db 781 AGTGCAGCTGGGCGATCCCGCCCGCGCGCTGGAAGAGAGGCGTGCAGCTGC 840  
Qy 998 TGCAGCTGGGCGAGCGCTTCTCAGCGTGCCTCGACGAGGACTTCCGCAAGTACACCG 1057  
Db 841 TGCAGCTGGGCGAGCGCTTCTCAGCGTGCCTCGACAGGACTTCCGCAAGTACACCG 900  
Qy 1058 CTTTACCATCCCGAGCATCAACAAGAGACCCCGCGCATCCGCTACAGTACAAACGTGC 1117  
Db 901 CTTTACCATCCCGAGCATCAACAAGAGACCCCGCGCATCCGCTACAGTACAAACGTGC 960  
Qy 1118 TGCAGCGCTGGAGGCGAGCGCCAGCATCTTCCAGAGCAGCATCAACAGAGCTGCG 1177  
Db 961 TGCAGCGCTGGAGGCGAGCGCCAGCATCTTCCAGAGCAGCATCAACAGAGCTGCG 1020  
Qy 1178 AGCCTTCCGCGCGCAACCCCGAGATCGTGATCTACAGTACATCGAGACCTGTACG 1237  
Db 1021 AGCCTTCCGCAAGCAGACCCCGCATCTGATCTACAGTACATCGACCTGTACG 1080  
Qy 1238 TGGGCAAGCGACCTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTGGCAAGCAC 1297  
Db 1081 TGGGCAAGCGACCTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTGGCGAGCAC 1140  
Qy 1298 TGTGCGCTGGGCTTCAACCCCGCAAGAAGCAGAGGAGCGCCCTTCTGT 1357  
Db 1141 TGTGCGCTGGGCTTCAACCCCGCAAGAAGCAGAGGAGCGCCCTTCTGT 1200  
Qy 1358 GGATGGCTACGAGCTGACCCCGCAAGTGGACCGTGCAGCGCATCGAGCTGCCCGAGA 1417  
Db 1201 GGATGGCTACGAGCTGACCCCGCAAGTGGACCGTGCAGCGCATCATGTGCCCGAGA 1260  
Qy 1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCGAGCC 1477  
Db 1261 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCGAGCC 1320  
Qy 1478 AGATCTACCCCGGATCAAGGTGGCGAGCTGTGCAAGCTGTGCGCGGCGCGAGGCGCC 1537  
Db 1321 AGATCTACCCCGGATCAAGGTGGCGAGCTGTGCAAGCTGTGCGCGGCGCGAGGCGCC 1380  
Qy 1538 TGACCGACATCTGCGGCTGACCGAGGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGA 1597  
Db 1381 TGACCGAGGTGATCCCTTACCGAGGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGA 1440

1598 TCCTGCGCAGCCGCTGACCGGTGTACTACACCCCAAGGACCTGGTGGCGAGA 1657  
 1441 TCCTGAAGAGCCCGTGCAGAGGTGTACTACGCCCAAGGACCTGGTGGCGAGA 1500  
 1658 TCAGAAGCAGGCGCCAGCAGCAGTGCAGCTTACAGATCTACAGGAGCCCTTCAAGAAC 1717  
 1501 TCAGAAGCAGGCGCCAGGCGAGTGCAGCTTACAGATCTACAGGAGCCCTTCAAGAAC 1560  
 1718 TGAAGACCGGCAAGTACGCAAGTGCAGCAGCCGCCACACCAACGACGCTGAAGCAGCTGA 1777  
 1561 TGAAGACCGGCAAGTACGCGCGATGCGCGCGGCCACACCAACGACGCTGAAGCAGCTGA 1620  
 1778 CCAGAGCCGTGCAGAGAGATCCCATCGAGAGCATCTGTGATCTGCGGCAAGACCCCAAGT 1837  
 1621 CCAGAGCCGTGCAGAGAGGTGAGCACCGAGAGCATCTGTGATCTGCGGCAAGATCCCAAGT 1680  
 1838 TCAGCCTGCCATCTCAGAGAGAGACCTGTGGAGACCTGTGGTGGACCGACTACTGTGAGGCCA 1997  
 1681 TCAAGCTGCCATCTCAGAGAGAGACCTGTGGAGAGCTGTGGTGGATGAGTACTGTGAGGCCA 1740  
 1898 CTTGGATCCCGAGTGGGAGTCTGTGAACACCCGCCCTGTGTGAAGCTGTGGTACCGC 1957  
 1741 CTTGGATCCCGAGTGGGAGTCTGTGAACACCCGCCCTGTGTGAAGCTGTGGTACCGC 1800  
 1958 TGAAGAGGAGCCCATCATCGCGCGCGAGACCTTTCTACGTGGAGCGCGCGCCCAACCGCG 2017  
 1801 TGAAGAGGAGCCCATCTGTGGCGCGGAGACCTTTCTACGTGGAGCGCGCGCCCAACCGCG 1860  
 2018 AGACCAAGATCGGCAAGCGCGCTAGCTGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 2077  
 1861 AGACCAAGCTGGGCAAGCGCGCTAGCTGACCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1920  
 2078 TGAACGAGACACCAAGAGAGCGAGTGTGAGCGAGCGAGCTGCGAGCGCGCGCGCGCGCGCG 2137  
 1921 TCAGCGACACCAAGAGAGCGAGTGTGAGCGAGCGAGCTGCGAGCGCGCGCGCGCGCGCGCG 1980  
 2138 GCGGCGAGCGAGTGAACATCTGACCGACGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCG 2197  
 1981 GCGGCGTGGAGGTGAACATCTGACCGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 2040  
 2198 AGCCCGACAAAGCGAGCGAGCTGTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 2257  
 2041 AGCCCGACAAAGCGAGCGAGCTGTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 2100  
 2258 AGAAGGTGTACCTGAGCTGGGTGCGCGCCACAGGCGATCGCGCGCGCGCGCGCGCGCGCG 2317  
 2101 AGAAGGTGTACCTGAGCTGGGTGCGCGCCACAGGCGATCGCGCGCGCGCGCGCGCGCGCG 2160  
 2318 ACAAGCTGTGAGCAAGGCGATCGCAAGGTGCTGTCTTCTGGAGCGGCGATCGATGGCGGCA 2377  
 2161 ACAAGCTGTGAGCGCGCGATCGCAAGGTGCTGTCTTCTGAGCGGCGATCGATGGCGGCA 2220  
 2378 TCGTGATCTACAGTATACGAGCAGCTGTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2437  
 2221 TCGTGATCTACAGTATACGAGCAGCTGTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2280  
 2438 AAAAGCTTCCGGGCTAGCACCGGTGAATTC 2469  
 2281 AAAAGCTTCCGGGCTAGCACCGGTGAATTC 2312

RESULT 6  
 AR373389  
 LOCUS AR373389 2312 bp DNA linear PAT 18-DEC-2003  
 DEFINITION Sequence 84 from patent US 6602705.  
 AR373389  
 ACCESSION AR373389  
 VERSION AR373389.1 GI:40075492  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE Unclassified.  
 1 (bases 1 to 2312)  
 AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.

TITLE Expression of HIV polypeptides and production of virus-like particles  
 JOURNAL Patent: US 6602705-A 84 05-AUG-2003;  
 FEATURES Location/Qualifiers  
 source 1..2312  
 /organism="unknown"  
 /mol\_type="genomic DNA"  
 ORIGIN  
 Query Match 83.1%; Score 2052; DB 6; Length 2312;  
 Best Local Similarity 93.8%; Pred. No. 9,6e-206;  
 Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
 170 GCGGCAAGGAGGCGCCACAGATGAAGGACTGTGACCGAGGCGCCAGGCGCAATCTTCTCGCG 229  
 1 GCGGCGCGCAAGGAGACCAAAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60  
 230 AGACCTGCGCTTCCCGCAGGCAAGCGCGCGAGTTCCCGCAGCAGAGAGAGCAACCGCGCA 289  
 61 AGACCTGCGCTTCTGCGAGGCAAGCGCGCGAGTTTCAGCAGCAGCAGCAGCAGCGCGCA 120  
 290 ACAGCGCCACCGCGCGAGCTGCGAGTGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 343  
 121 ACAGCGCCACCGCGCGAGCTGCGAGTGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
 344 CCGCGCGCGCGCGCGAGCG 397  
 181 CCGCGCGCGCGCGCGAGCG 240  
 398 GCGCGCTGTGAGCATCAAGGTGGCGCGCGAGTCAAGAGGCGCGCGCGCGCGCGCGCGCG 457  
 241 GCGCGCTGTGAGCATCAAGGTGGCGCGCGAGTCAAGAGGCGCGCGCGCGCGCGCGCGCG 300  
 458 CGACGACCGCGCGTGTGAGGAGTCAAGGTGGCGCGCGAGTCAAGAGGCGCGCGCGCGCGCG 517  
 301 CGACGACCGCGCGTGTGAGGAGTCAAGGTGGCGCGCGAGTCAAGAGGCGCGCGCGCGCG 360  
 518 GCGCGCATCGCGCGCTTCAAGGTGGCGCGAGTCAAGGTGGCGCGAGTCAAGGTGGCGCG 577  
 361 GCGCGATCGCGCGCTTCAAGGTGGCGCGAGTCAAGGTGGCGCGAGTCAAGGTGGCGCG 420  
 578 GCAAGAGCGCATCGCGCGCGTGTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637  
 421 GCGCAAGGCGCATCGCGCGCGTGTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 480  
 638 ACATGTGACCGCGCGTGTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 697  
 481 ACCTGTGACCGCGCGTGTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
 698 CCGTGAAGCTGAAGCG 757  
 541 CCGTGAAGCTGAAGCG 600  
 758 AGAAGATCAAGCA 817  
 601 AGAAGATCAAGCA 660  
 818 AGATCGCGCGCGGAGAGCG 877  
 661 AGATCGCGCGCGGAGAGCG 720  
 878 CCAAGTGGCGCGCAAGCTGTGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 937  
 721 CCAAGTGGCGCGCAAGCTGTGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
 938 AGGTGAGCTGTGGCG 997  
 781 AGGTGAGCTGTGGCG 840  
 998 TGGACGTGGCGCGAGCGCGCTACTTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057  
 841 TGGACGTGGCGCGAGCGCGCTACTTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900



```
Qy 1058 CCTTCAACATCCCGAGCATCAACACGAGAGCCCGGGATCCGCTACAGTACAAACGTGC 1117
Db 901 CCTTCAACATCCCGAGCATCAACACGAGAGCCCGGGATCCGCTACAGTACAAACGTGC 960
Qy 1118 TGCCCCAGGGCTGGAAGGGAGCCCGAGCATCTTCCAGAGCAGCATCAACAAATCTCTGG 1177
Db 961 TGCCCCAGGGCTGGAAGGGAGCCCGGGATCTTCCAGAGCAGCATCAACAAATCTCTGG 1020
Qy 1178 AGCCCTTCCGGCCCGCAACCCCGAGATTCGTGATCTTACCAATACATGGAACCTGTACG 1237
Db 1021 AGCCCTTCCGGCAAGACAGAACCCCGACATCGTGATCTTACCAATACATGGAACCTGTACG 1080
Qy 1238 TGCGCAGGACCTTGGAGATCGGCCAGCACCAGCCGCAAGATCGAGAGCTGGCAAGCACC 1297
Db 1081 TGCGCAGGACCTTGGAGATCGGCCAGCACCAGCCGCAAGATCGAGAGCTGGCGCAGCACC 1140
Qy 1298 TGCTGCGCTGGGGCTTCAACACCCCGCAAGAGTGGACCGTGCAGCCCATCGAGCTGCCGAGA 1417
Db 1141 TGCTGCGCTGGGGCTTCAACACCCCGCAAGAGTGGACCGTGCAGCCCATCGAGCTGCCGAGA 1200
Qy 1358 GGTATGGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGA 1417
Db 1201 GGTATGGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGA 1260
Qy 1418 AGGAGACTGGACCGTGAAACGATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGCAGCC 1477
Db 1261 AGGAGACTGGACCGTGAAACGATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGCAGCC 1320
Qy 1478 AGATCTACCCCGGATCAAGGTGGCGGAGCTGTCAAGCTGTCTGCGGGCGCCAAAGGCC 1537
Db 1321 AGATCTACCCCGGATCAAGGTGGCGGAGCTGTCAAGCTGTCTGCGGGCGCCAAAGGCC 1380
Qy 1538 TGACCGACATCGTGCCCTGACCCGAGGAGCGCAGCTGGAGCTGGCGGAGAACCGCGAGA 1597
Db 1381 TGACCGAGGTGATCCCTTGACCGAGGAGCGCAGCTGGAGCTGGCGGAGAACCGCGAGA 1440
Qy 1598 TCCTGCGCGAGCCCGTGACGCGGTGTACTAGACCCCGAGAGGACTGGTGGCGCGAGA 1657
Db 1441 TCCTGAGAGGAGCCCGTGACGAGGTGTACTAGACCCCGAGAGGACTGGTGGCGCGAGA 1500
Qy 1658 TCAGAGAGCGGCGCAGACCGAGTGGACCTACAGATCTACAGAGCCCTTCAAGAAC 1717
Db 1501 TCAGAGAGCGGCGCAGGCGGCGAGTGGACCTACAGATCTACAGAGCCCTTCAAGAAC 1560
Qy 1718 TGAAGACCGGCAAGTACGCGCAAGTGGCGACCGCGCCACACCAAGAGTGAAGCAGCTGA 1777
Db 1561 TGAAGACCGGCAAGTACGCGCGATGCGCGGCGCCACACCAAGAGTGAAGCAGCTGA 1620
Qy 1778 CCGAGGCGGTGCAGAAAGATCGCCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGT 1837
Db 1621 CCGAGGCGGTGCAGAAAGTGCAGACCGAGAGCATCTGTGATCTGGGGCAAGATCCCAAGT 1680
Qy 1838 TCCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGTGTGACCGGACTCTGGCAGGSCCA 1897
Db 1681 TCAAGCTGCGCCATCCAGAGGAGACCTGGGAGGCTGTGTGATGTGAGTACTGGCAGGSCCA 1740
Qy 1898 CTTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCTGGTGAAGCTGTGTGATCCAGC 1957
Db 1741 CTTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCTGGTGAAGCTGTGTGATCCAGC 1800
Qy 1958 TGAGAAAGGAGCCCATCATCGGCGCCGAGACCTTCTAAGTGGAGCGGCGCCGCAACCGCG 2017
Db 1801 TGAGAAAGGAGCCCATCATCGTGGCGCCGAGACCTTCTAAGTGGAGCGGCGCCGCAACCGCG 1860
Qy 2018 AGACCAAGATCGCAAGGCGGCTACGTGACCGACCGGGCGCGGCGGCAAGATCGTGAGCC 2077
Db 1861 AGACCAAGCTGGGCAAGGCGGCTACGTGACCGACCGGGCGCGGCGGCAAGATCGTGAGCA 1920
Qy 2078 TGACCGAGACACCAACAGAGACCGAGCTGACAGGCGCATCCAGCTGGCTGTCAGGACA 2137
Db 1921 TCGCCGACACACCAACAGAGACCGAGCTGACAGGCGCATCCACCTGGCTGTCAGGACA 1980
Qy 2138 GCGGACGAGGAGTGAACATCTGTGACCGAGACCGCATCGCCCTGGGGCATCATCCAGGCC 2197
```

```
Db 1981 GCGGCTTGGAGGTGAACATCGTGACCCAGACAGCCAGTAGCCCTGGGCAATCATCCAGGCC 2040
Qy 2198 AGCCCGCAAGAGAGCGAGAGGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGG 2257
Db 2041 AGCCCGCAAGAGAGCGAGAGGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGG 2100
Qy 2258 AGAAGGTGTACTAGAGCTGGGTGCCCGCCCAAGAGGAGTCCGGCGGCAACAGAGCATCG 2317
Db 2101 AGAAGGTGTACTAGAGCTGGGTGCCCGCCCAAGAGGAGTCCGGCGGCAACAGAGCATGG 2160
Qy 2318 ACAGAGCTGGTGAAGAGGAGTCCGCAAGGTGCTGTTCTTGAACGGCATCGATGGCGGCA 2377
Db 2161 ACAGAGCTGGTGAAGAGGAGTCCGCAAGGTGCTGTTCTTGAACGGCATCGATGGCGGCA 2220
Qy 2378 TCCTGATCTACCAAGTATCATGACACCTGTACGTGGGAGCGGCGGCTTAGGATCGATT 2437
Db 2221 TCCTGATCTACCAAGTATCATGACACCTGTACGTGGGAGCGGCGGCTTAGGATCGATT 2280
Qy 2438 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2469
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 7
AX427930          AX427930          9166 bp      DNA      linear      PAT 20-JUN-2002
DEFINITION       Sequence 168 from Patent WO0232943.
ACCESSION        AX427930
VERSION          AX427930.1 GI:21538017
KEYWORDS         .
SOURCE           synthetic construct
ORGANISM         synthetic construct
REFERENCE        1
AUTHORS          Huang, Y. and Nabel, G.J.
TITLE            Modifications of hiv env, gag, and pol enhance immunogenicity for
                genetic immunization
                Patent: WO 0232943-A 168 25-APR-2002;
                GOVERNMENT OF THE UNITED STATES (US)
JOURNAL
FEATURES         Location/Qualifiers
                source
                1..9166
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match      83.1%; Score 2051.2; DB 6; Length 9166;
Best Local Similarity 92.6%; Pred. No. 8.2e-206;
Matches 2192; Conservative 0; Mismatches 163; Indels 13; Gaps 3;

Qy 14 TGCGCGAGGCGCATGAGCCAGGCCACAGCGCCACATCTCTGATGCGAGCGCAACTTCA 73
Db 2961 TGGCGGAGGCGCATGAGCAAGGTGAACCAACCAACCATCATGATGCGAGCGCAACTTCA 3020
Qy 74 AGGCGCCCAAGAGCGCATCATCAAGTGTCTCAACTCGCGCAAGAGGCGGCACATCGCCGCA 133
Db 3021 AGGCGCCCAAGAGCGCATCATCAAGTGTCTCAACTCGCGCAAGAGGCGGCACATCGCCGCA 3080
Qy 134 ACTGCGCGCCCGCCGCAAGAGGCTGTGTGAAGTGTGGCAAGAGGCGGCACACAGATGA 193
Db 3081 ACTGCGCGCCCGCCGCAAGAGGCTGTGTGAAGTGTGGCAAGAGGCGGCACACAGATGA 3140
Qy 194 AGGACTGCAAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 253
Db 3141 AGGACTGCAAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 3199
Qy 254 AGGCGCGCGGAGTTTCCCGAGCGAGCAGAAACCGCGCCAAACAGCCCCCAGCGCGGCTGC 313
Db 3200 AGGCGCGGAGATTTTCTTCTAGAGCAGACAGAGCCACAGCCCAACAGAGAGAGCTTC 3259
Qy 314 AGGTGCGCGG-----CGAACACCCCGCGAGCGAGGCGCGGCGGCGGCGGCGGCGGCGGCA--- 364
```

Db 3260 AGGTTTGGGGAAGAGCAACAACTCTCTCAGAAAGCAGGAGCCGATAGACAAGGAACGTG 3319  
 Qy 365 ---CCCTGAATCTCCCCAGATCAACCTGTGGCAGCGCCCTTGGTGAGGATCAAGGTGG 421  
 Db 3320 TATCTTTAGCTTCTCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAACAATAAGATAG 3379  
 Qy 422 GCGGCCAGATCAAGGAGGCGCTCTGGAACACCGCGCCGAGACACCGTGTCTGGAGGAGA 481  
 Db 3380 GGCGCCAGCTGAAGGAGGCGCTTCTAGACACCGCGCCGAGACACCGTGTCTGGAGGAGA 3439  
 Qy 482 TGAGCCTTGCCCGGCAAGTGGAAGCCCAAGATGATCGCGGCATCGCGGCTTCAATCAAGG 541  
 Db 3440 TGAACCTGCGCGCGCTGGAAGCCCAAGATGATCGCGGCATCGCGGCTTCAATCAAGG 3499  
 Qy 542 TGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGCGCAAGAAGGCGCATCGGACCGTGC 601  
 Db 3500 TGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGCGCAAGAAGGCGCATCGGACCGTGC 3559  
 Qy 602 TGATCGGCCCGCCACCCCGTGAAATCATCGCGCGCAACATGCTGACCCAGCTCGGCTGCA 661  
 Db 3560 TGGTGGGCCCGCCACCCCGTGAAATCATCGCGCGCAACATGCTGACCCAGATCGGCTGCA 3619  
 Qy 662 CCCTGAACTTTCCCATCAGCCCATCGAGACCGTGGCCGCTGGAAGCTGAAGCCCGGCATGG 721  
 Db 3620 CCCTGAACTTTCCCATCAGCCCATCGAGACCGTGGCCGCTGGAAGCTGAAGCCCGGCATGG 3679  
 Qy 722 ACGGCCCAAGGTGAAGCAGTGGCCCTTGACCCAGAGAGAAGATCAAGGCCCTGACCGCCA 781  
 Db 3680 ACGGCCCAAGGTGAAGCAGTGGCCCTTGACCCAGAGAGAAGATCAAGGCCCTGAGGAGA 3739  
 Qy 782 TCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACAAATCGCGCCCGGAGAACCCCTACA 841  
 Db 3740 TCTGCAACCGAGATGGAAGAGGAGGCAAGATCAAGAAAGATCGGCCCCGAGAACCCCTACA 3799  
 Qy 842 ACACCCCGCTTTCGCCATCAAGAAGAGGAGCAAGCAAGTGGCGCAAGCTGGTGAGT 901  
 Db 3800 ACACCCCGCTTTCGCCATCAAGAAGAGGAGCAAGCAAGTGGCGCAAGCTGGTGAGT 3859  
 Qy 902 TCGCGAGCTGAACAGCGCACCCAGATCTTCTGGAGGTGAGCTGGGGATCCCCCACC 961  
 Db 3860 TCGCGAGCTGAACAGCGCACCCAGATCTTCTGGAGGTGAGCTGGGGATCCCCCACC 3919  
 Qy 962 CCGCCGCGCTGAAGAGAGAGAGAGAGTGGAGCTGGAGCTGGGCGAGCGCTACTTCA 1021  
 Db 3920 CCGCCGCGCTGAAGAGAGAGAGAGTGGAGCTGGAGCTGGGCGAGCGCTACTTCA 3979  
 Qy 1022 GCGTGGCCCTTGAGCAGGAGCTTCCGCAAGTACACCCGCTTTCACCATCCCCAGCATCAACA 1081  
 Db 3980 GCGTGGCCCTTGAGCAGGAGCTTCCGCAAGTACACCCGCTTTCACCATCCCCAGCATCAACA 4039  
 Qy 1082 ACAGAACCCCGGATCCGCTACAGTAAAGTGTGCCCCAGGGCTGGAAGGGCAGCC 1141  
 Db 4040 ACAGAACCCCGGATCCGCTACAGTAAAGTGTGCCCCAGGGCTGGAAGGGCAGCC 4099  
 Qy 1142 CCAGCATCTTCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201  
 Db 4100 CCGCCATCTTCAGAGTGAAGATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCG 4159  
 Qy 1202 AGATCGTGATCTACAGTACATGAGCAGCTGTACGTGGGCGAGCAGCTTGGAGATCGGCC 1261  
 Db 4160 ACATCGTGATCTACAGTACATGAGCAGCTGTACGTGGGCGAGCAGCTTGGAGATCGGCC 4219  
 Qy 1262 AGACCGCGCAAGATCGAGGAGTGGCGAAGCAACCTGCTGGCTGGGGCTTCAACACCC 1321  
 Db 4220 AGACCGCGCAAGATCGAGGAGTGGCGAAGCAACCTGCTGGCTGGGGCTTCAACACCC 4279  
 Qy 1322 CCAGCAAGAGACCAAGAGCGCCCTTCTCTGTGATGGGCTACGAGCTGCACCCCG 1381  
 Db 4280 CCAGCAAGAGACCAAGAGCGCCCTTCTCTGTGATGGGCTACGAGCTGCACCCCG 4339  
 Qy 1382 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACA 1441

Db 4340 ACAAGTGGACCGTGCAGACCCATCGTGTGCTGCCGAGAAAGGACAGCTGGACCGTGAACGACA 4399  
 Qy 1442 TCCAGAAAGCTGGTGGCAAGCTGAACCTGGGCGCAGCCAGATCTTACCCTGGCATCAAGGTGC 1501  
 Db 4400 TCCAGAAAGCTGGTGGCAAGCTGAACCTGGGCGCAGCCAGATCTTACGCGCGCATCAAGGTGC 4459  
 Qy 1502 GCCAGCTGTCAAGCTGTGCGCGCGCCAAAGCCCTTGAACATCGTGTGCCCTGACCG 1561  
 Db 4460 GCCAGCTGTCAAGCTGTGCGCGCGCCAAAGCCCTTGAACATCGTGTGCCCTGACCG 4519  
 Qy 1562 AGAGGCCAGAGCTGGAGCTGGCGCGAAGCCGCGAGATCTTCCGCGAGCCCGTGCACGGCG 1621  
 Db 4520 AGAGGCCAGAGCTGGAGCTGGCGCGAAGCCGCGAGATCTTGAAGSAGCCCGTGCACGGCG 4579  
 Qy 1622 TGTACTACGACCCCAAGAGCTGTGCGCGAGATCCAGAAAGCAGGCGCACGACAGT 1681  
 Db 4580 TGTACTACGACCCCAAGAGCTGTGCGCGAGATCCAGAAAGCAGGCGCACGAGT 4639  
 Qy 1682 GGAACCTTACAGATCTTACCAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGA 1741  
 Db 4640 GGAACCTTACAGATCTTACCAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCGCA 4699  
 Qy 1742 TGCGCACCGCCCAACAGAGCTGAAGAGCTGACCGAGGCGCTGCAGAAAGATCGCCA 1801  
 Db 4700 TGAAGGGCGGCCCAACAGAGCTGAAGAGCTGACCGAGGCGCTGCAGAAAGATCGCCA 4759  
 Qy 1802 TGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTTCCGCTTCCCATCCAGAAAGAGA 1861  
 Db 4760 CCGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTTCCGCTTCCCATCCAGAAAGAGA 4819  
 Qy 1862 CTTGGGAGACCTTGGTGGACCGACTACTTGGCAGGCGACCTTGGATCTCCGAGTGGGAGTTCG 1921  
 Db 4820 CTTGGGAGGCTTGGTGGACCGAGTACTTGGCAGGCGACCTTGGATCTCCGAGTGGGAGTTCG 4879  
 Qy 1922 TGAACACCCCGCTTGGTGGAGCTGTGATCCAGCTGGAAGAGGAGCCCATCATCGGCG 1981  
 Db 4880 TGAACACCCCGCTTGGTGGAGCTGTGATCCAGCTGGAAGAGGAGCCCATCATCGGCG 4939  
 Qy 1982 CCGAGACCTTCTACGTGGAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCT 2041  
 Db 4940 CCGAGACCTTCTACGTGGAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCT 4999  
 Qy 2042 ACGTGACCGACCGGGCGCGCGCAAGATCGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 2101  
 Db 5000 ACGTGACCGACCGCGCGCGCGCAAGATCGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 5059  
 Qy 2102 CCGAGCTGAGGCGCATCGAGCTGGCGCTGAGAGACGCGCGAGCGAGGTGAACATCGTGA 2161  
 Db 5060 CCGAGCTGAGGCGCATCGAGCTGGCGCTGAGAGACGCGCGAGCGAGGTGAACATCGTGA 5119  
 Qy 2162 CCGACAGCGAGTACGCGCTGGGCGATCATCCAGGCGCGAGCGCGCAAGAGCGAGAGCGAGC 2221  
 Db 5120 CCGACAGCGAGTACGCGCTGGGCGATCATCCAGGCGCGAGCGCGCAAGAGCGAGAGCGAGC 5179  
 Qy 2222 TGGTGAACCGAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACCTGAGCTGGGTGC 2281  
 Db 5180 TGGTGAACCGAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACCTGAGCTGGGTGC 5239  
 Qy 2282 CCGCCCAAGAGGCGCATCGCGCGCAACGAGCAGATCCGACAGCTTGGTGAAGAGGCGATCC 2341  
 Db 5240 CCGCCCAAGAGGCGCATCGCGCGCAACGAGCAGGTGGAAGCGGCTGGTGAAGCGCGGATCC 5299  
 Qy 2342 GCAAGGTGTCTTCTTGGAGCGGATCGA 2369  
 Db 5300 GCAAGGTGTCTTCTTGGAGCGGATCGA 5327

RESULT 8  
 AX427936  
 LOCUS 9788 bp DNA linear PAT 20-JUN-2002  
 DEFINITION Sequence 174 from Patent WO232943.  
 AX427936  
 ACCESSION AX427936.1 GI:21538023  
 VERSION



```
QY 1862 CTTGGGAGACCTTGGTGGACCGAAGTCTACTGGCAGGCCACCTTGGATCCCGAGTGGGAGTTGG 1921
Db 4815 CTTGGGAGGCTTGGTGGACCGAGTACTTGGCAGGCCACCTTGGATCCCGAGTGGGAGTTGG 4874
QY 1922 TGAACACCCCCCTTGGTGAAGCTGTGTGTAACAGCTTGGAGAGGAGGCCATCATCGGGG 1981
Db 4875 TGAACACCCCCCTTGGTGAAGCTGTGTGTAACAGCTTGGAGAGGAGGCCATCATCGGGG 4934
QY 1982 CCGAGACCTTCTACTGTGGAGCGGCCCGCCACCGAGACCGACCAAGATCGGGCAGGCCGCT 2041
Db 4935 CCGAGACCTTCTACTGTGGAGCGGCCCGCCACCGAGACCGACCAAGCTTGGGCAAGGCCGCT 4994
QY 2042 ACCTGACCGACCGGGCGCGCAGAGATCGTGAAGCTGACCGGAGACCAACCAAGCAAGA 2101
Db 4995 ACCTGACCGACCGGGCGCGCAGAGATCGTGAAGCTGACCGGAGACCAACCAAGCAAGA 5054
QY 2102 CCGAGCTGCAAGGCCATCCAGCTGGCCCTGCAAGACAGCGGACCGAGGTGAACATCGTGA 2161
Db 5055 CCGAGCTGCAAGGCCATCCAGCTGGCCCTGCAAGACAGCGGACCGAGGTGAACATCGTGA 5114
QY 2162 CCGAGACCGAGTACGCCCTGGGCTGATCATCCAGGCCCGACCGCCGACCAAGAGGAGCGAGC 2221
Db 5115 CCGAGACCGAGTACGCCCTGGGCTGATCATCCAGGCCCGACCGCCGACCAAGAGGAGCGAGC 5174
QY 2222 TGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGC 2281
Db 5175 TGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGC 5234
QY 2282 CCGCCCAAGGGCATCGGGCGGCAACGAGCAGATCGACCAAGCTGGTGAAGCAAGGGCATCC 2341
Db 5235 CCGCCCAAGGGCATCGGGCGGCAACGAGCAGGTGAGCGGCTGGTGAAGCGCGGCATCC 5294
QY 2342 GCAAGTGTCTTCTTGGAGCGGATCGATGGCGGCTCGTGATCTACCACTA 2393
Db 5295 GCAAGTGTCTTCTTGGAGCGGATCGACAAAGGCCCGAGGAGGACGAGAA 5346

RESULT 9
AX427931
LOCUS AX427931 9169 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 169 from Patent WO232943.
ACCESSION AX427931
VERSION AX427931.1 GI:21538018
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Huang, Y. and Nabel, G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization
JOURNAL Patent: WO 0232943-A 169 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
    Location/Qualifiers
        1..9169
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"
            /notes="plasmid pVR1012x/s containing HIV genes"
ORIGIN
    Query Match 82.2%; Score 2028.8; DB 6; Length 9169;
    Best Local Similarity 91.7%; Pred. No. 1.8e-203;
    Matches 2194; Conservative 0; Mismatches 182; Indels 16; Gaps 4;

QY 14 TGGCCGAGGCGCATGAGCCAGGCCACCAGCGGCCAACATCTCTGATGCGAGCGCAGCAACTTCA 73
Db 2967 TGGCCGAGGCGCATGAGCCAGGTCGAGCAGACACACATCATGATGCGAGCGGCAACTTCA 3026
QY 74 AGGGCCCCAAGCGGCATCATCAAGTGTTCAACTTGGCGGCAAGGAGGGGCACATCGCCGCA 133
Db 3027 AGGGCCCCAAGCGGCAT---CAAGTGTTCAACTTGGCGGCAAGGAGGGGCACATCGCCGCA 3083
```

```
QY 134 ACTGCCGCGCCCCCGCAAGAAAGGGCTGCTGGAAGTGCGBGCAAGGAGGCGCACAGATGA 193
Db 3084 ACTGCCGCGCCCCCGCAAGAAAGGGCTGCTGGAAGTGCGBGCAAGGAGGCGCACAGATGA 3143
QY 194 AGGACTGCAACGAGGCCAGGGCCAACTTCTTCGCGAGGACCTTGGGCTTCCCCGAGGGCA 253
Db 3144 AGGACTGCAACGAGGCCAGGGCTAA--TTTTTTAGGGAAGATCTGGGCTTCCCAAGAGGA 3202
QY 254 AGGCCCGCGAGTTCCCGAGGAGCAAGCGCGCCAAAGCCAGCCACCAGCGCGAGCTGC 313
Db 3203 AGGCCGAGGAAATTTTCTTTCAGAGCAGACAGAGCCAAAGCCCAACAGAGAGGCTTC 3262
QY 314 AGGTGCGCGG-----CGACAAACCCCGCAGCGAGGCCGCGCGCGCAGCGCAGGGCA-- 364
Db 3263 AGGTTTGGGGAAGAGACAACACTCCCTCTCAGAAGCAGGAGCCGATAGACAGGAACCTG 3322
QY 365 ---CCCTGAACCTTCCCCAGATCACCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGGTGG 421
Db 3323 TATCTCTTAGCTTCCCTCAGATCACTCTTTTGGCAGCGAGCCCTCGTCAAAATAAGATAG 3382
QY 422 GCGGCCAGATCAGGAGGCGCTCTGGAGACACCGCGCGCGCAGACACCGTGTCTGGAGGAGA 481
Db 3383 GGGGCCAGCTGAAGAGGAGGCCCTTCTAGACACCGCGCGCCGACGACACCGTGTCTGGAGGA 3442
QY 482 TGAGCTGCTCCCGCAAGTGGAAAGCCCAAGATGATCGCGCGCGCATCGGCGGCTTCAAGG 541
Db 3443 TGAACCTGCGCGCGCTTGGAGGCCCAAGATGATCGCGCGCATCGGCGGCTTCAAGG 3502
QY 542 TGGCCGAGTACGACACAGATCTGTATCGAGATCTTGGGCAAGAGGCCATCGGCACCGTGC 601
Db 3503 TGGCCGAGTACGACACAGATCTGTATCGAGATCTTGGCGGCCACAAAGGCCATCGGCACCGTGC 3562
QY 602 TGATCGGCGCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTCGGCTGCA 661
Db 3563 TGGTGGGCGCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGATCGGCTGCA 3622
QY 662 CCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGAAAGCCCGCATGG 721
Db 3623 CCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGAAAGCCCGCATGG 3682
QY 722 AGGCCCGCAAGGTGAAGCAGTGCGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCA 781
Db 3683 ACGGCCCGCAAGGTGAAGCAGTGCGCCCTGACCGAGGAGAGATCAAGGCCCTGCGTGAGA 3742
QY 782 TCTGCGAGGAGATGGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACA 841
Db 3743 TCTGCAACCGAGATGGAGAGAGGAGGCAAGATCAGCAAGATCGGCCCGGAGAACCCCTACA 3802
QY 842 ACACCCCGCTTTCGCCCATCAAGAAGAGGACAGACCAAGTGGCGCAAGCTGGTGGACT 901
Db 3803 ACACCCCGCTTTCGCCCATCAAGAAGAGGACAGACCAAGTGGCGCAAGCTGGTGGACT 3862
QY 902 TCGCGAGCTGAACAAGCGGACCCAGGACTTCTGGGAGGTGACGCTGGGGCATCCCCACC 961
Db 3863 TCGCGAGCTGAACAAGCGGACCCAGGACTTCTGGGAGGTGACGCTGGGGCATCCCCACC 3922
QY 962 CCGCGCGCTGAGAGAGAGAGAGCGTGCACCGTGTGGAAGTGGCGAGCGCTACTTCA 1021
Db 3923 CCGCGCGCTGAGAGAGAGAGAGCGTGCACCGTGTGGAAGTGGCGAGCGCTACTTCA 3982
QY 1022 GCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTACCCTCCCGAGCATCAACA 1081
Db 3983 GCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTACCCTCCCGAGCATCAACA 4042
QY 1082 ACAGAACCCCGGCGATCCGCTACAGTACAAAGTGTGCGCCGAGGGCTGGAAGGGCAGCC 1141
Db 4043 ACAGAACCCCGGCGATCCGCTACAGTACAAAGTGTGCGCCGAGGGCTGGAAGGGCAGCC 4102
QY 1142 CGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCCGCGCCCGCAACCCCG 1201
Db 4103 CGCCCATCTTCCAGTGCAGATGACCAAGATCTCTGGAGCCCTTCCCGAGCAGACACCCCG 4162
QY 1202 AGATCGTGATCTACCAAGTACATGGACGACCTGTACGTGGGCGAGCGACCTGGAGATCGGCC 1261
```

Db 4163 ACATCGTGATCTACAGTACATGGACCACTGTATCGTGGGAGCGACCTGGAGATCGGCC 4222  
Qy 1262 AGCACCGCGCCAAAGATCGAGAGCTGCGCAAGCACCTCTGCGCTGGGGCTTCCACCAACC 1321  
Db 4223 AGCACCGCACCAAGATCGAGAGCTGCGCCAGACCTCTGCGCTGGGGCTTCCACCAACC 4282  
Qy 1322 CCAGACAGAACACACAGAGAGAGCCCTTCTCTGTGGATGGGTACGAGCTGCACCCCG 1381  
Db 4283 CCAGACAGAACACACAGAGAGAGCCCTTCTCTGTGGATGGGTACGAGCTGCACCCCG 4342  
Qy 1382 ACAGTGCAGCGTCAGCCATCGAGCTGCGCGAGAGAGAGCTGGACCTGGAACGACA 1441  
Db 4343 ACAAGTGCAGCGTCAGCCATCGAGCTGCGCGAGAGAGAGCTGGACCTGGAACGACA 4402  
Qy 1442 TCAGAAAGCTGGTGGGCAAGCTGAACCTGGGCGAGCCAGATCTACCCCGCATCAAGGTGC 1501  
Db 4403 TCAGAAAGCTGGTGGGCAAGCTGAACCTGGGCGAGCCAGATCTACCCCGCATCAAGGTGC 4462  
Qy 1502 GCCAGCTGTGAAGCTGTGCGCGGCGCAAGGCTCTGACCGACATCGTGCCTTGACCG 1561  
Db 4463 GCCAGCTGTGAAGCTGTGCGCGGCGCAAGGCTCTGACCGAGCTGTGCTGCCCTGACCG 4522  
Qy 1562 AGGAGCGGAGCTGGAGCTGGCGGAGACCGGAGATCTGCGGAGCCCTGCGACCGCG 1621  
Db 4523 AGGAGCGGAGCTGGAGCTGGCGGAGACCGGAGATCTGCGGAGCCCTGCGACCGCG 4582  
Qy 1622 TGTACTACGACCCAGCAAGACCTGTGTGGCGGAGATCTGCGGAGCCCTGCGACCGAGT 1681  
Db 4583 TGTACTACGACCCAGCAAGACCTGTGTGGCGGAGATCTGCGGAGCCCTGCGACCGAGT 4642  
Qy 1682 GGACTTACAGATCTACAGAGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAGA 1741  
Db 4643 GGACTTACAGATCTACAGAGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCA 4702  
Qy 1742 TGCGCACCGCCACACCAACGAGCTGAGAGCTGACCGAGGCGCTGCGAGAGATCGCCA 1801  
Db 4703 TGAAGGCGCGCCACACCAACGAGCTGAGAGCTGACCGAGGCGCTGCGAGAGATCGCCA 4762  
Qy 1802 TGAGAGAGATCGTGATCTGGGGCAAGACCCCAAGTCTCCGCTGCCATCCAGAGGAGA 1861  
Db 4763 CCAGAGAGATCGTGATCTGGGGCAAGACCCCAAGTCTCCGCTGCCATCCAGAGGAGA 4822  
Qy 1862 CTTGGGAGACTGTGTGACCGACTACTGGCAGGCCACTTGGATCCCGAGTGGAGTTGC 1921  
Db 4823 CTTGGGAGGCTGTGTGACCGACTACTGGCAGGCCACTTGGATCCCGAGTGGAGTTGC 4882  
Qy 1922 TGAACACCCCGCTCGTGTGAGCTGTGTGACAGCTGAGAGAGGAGCCATCATCGCG 1981  
Db 4883 TGAACACCCCGCTCGTGTGAGCTGTGTGACAGCTGAGAGAGGAGCCATCATCGCG 4942  
Qy 1982 CCAGAGCTTCTAGCTGAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCT 2041  
Db 4943 CCAGAGCTTCTAGCTGAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCT 5002  
Qy 2042 ACGTGACCGACCGGGCGCGCAGAGATCGTGAGCTTGAACCGAGACCAACACAGAGA 2101  
Db 5003 ACGTGAACCGACCGCGCGCGCAGAGGCTGTGTGCGCTGACCGACCAACACAGAGA 5062  
Qy 2102 CCAGAGCTGAGGCGCATTCAGCTGCGCTGAGGACAGCGGAGGAGGTGAACATCGTGA 2161  
Db 5063 CCAGAGCTGAGGCGCATTCAGCTGCGCTGAGGACAGCGGCGCTGAGGAGTGAACATCGTGA 5122  
Qy 2162 CCAGACGCCAGTACCGCTGGGCATCTCCAGGCGCCAGCGGCGCAAGAGCGAGAGCGAGC 2221  
Db 5123 CCAGACGCCAGTACCGCTGGGCATCTCCAGGCGCCAGCGGCGCAAGAGCGAGAGCGAGC 5182  
Qy 2222 TGTGAACCCAGATCATCGAGCAGCTGATCAAGAGAGGAGAGTGTACTGAGCTGGGTGC 2281  
Db 5183 TGTGAGCCAGATCATCGAGCAGCTGATCAAGAGAGGAGAGTGTACTGAGCTGGGTGC 5242  
Qy 2282 CCGCCCAACAGGGGATCGGGCGAACCGAGCAGATCTGAGCAGCTGGTGGAGAGGCGATCC 2341

Db 5243 CCGCCCAACAGGGGATCGGGCGCAACGAGAGGTGGACGGCTGTGAGCGCGCGATCC 5302  
Qy 2342 GCAAGGTGCTCTTCTGAGACGGCATCGATGCGGCGCATCTGATCTTACAGTA 2393  
Db 5303 GCAAGGTGCTCTTCTGAGACGGCATCGACAAAGGCCAGAGGAGCACGAGAA 5354  
  
RESULT 10  
LOCUS AX427925 9194 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 163 from Patent WO0232943.  
ACCESSION AX427925  
VERSION AX427925.1 GI:21538012  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
  
REFERENCE 1  
AUTHORS Huang,Y. and Nabel,G.J.  
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization  
JOURNAL Patent: WO 0232943-A 163 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
  
FEATURES  
source Location/Qualifiers  
1..9194  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"  
  
ORIGIN  
  
Query Match 82.1%; Score 2027; DB 6; Length 9194;  
Best Local Similarity 91.6%; Pred. No. 2.8e-203;  
Matches 2194; Conservative 0; Mismatches 185; Indels 16; Gaps 4;  
  
Qy 14 TGGCCGAGGCGCATGAGCCAGG---CCACCAGCGCCACATCTGATGAGCGCAGCAACT 70  
Db 2967 TGGCCGAGGCGCATGAGCGAGGTGACCAACAGCGCCACCATGATGAGCGCGCAACT 3026  
Qy 71 TCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGCGCAAGAGAGGCGCCACATCGCCC 130  
Db 3027 TCCGCAACCGAGCGCAAGATCGTGAAGTCTTCAACTGCGGCAAGAGAGGCGCCACCGCCC 3086  
Qy 131 GCAACTCCGCGCCCCCGCGCAAGAGGGCTGTGGAAGTGCAGAGAGAGGCGCCACCAAGA 190  
Db 3087 GCAACTCCGCGCCCCCGCGCAAGAGGGCTGTGGAAGTGCAGAGAGGCGCCACCAAGA 3146  
Qy 191 TGAAGGACTGACCGAGGCGCCAGCCCAACTTCTTCCGCGAGGACCTGCGCTTCCCGCAGG 250  
Db 3147 TGAAGGACTGACCGAGGCGCAGGCTAA-TTTTTTTGAAGGAAGATCTGCGCTTCCCGCAGG 3205  
Qy 251 GCAAGGCGCGGAGTTCCTCCAGCGAGCAAGAACCGCGCCACAGCCCAACAGCGCGCGAGC 310  
Db 3206 GGAAGGCGCAGGGAATTTCTTCAGAGCAGACCAAGCCACAGCCCAACAGAGAGAGC 3265  
Qy 311 TGCAAGTGCAGCGG-----CGACAAACCCCGCAGCGAGGCGCGCGCGCGCCAGGCGCA 364  
Db 3266 TTCAGGTTTGGGGAAGAGACAACTCCCTCTCAGAAGCAGGAGCGCGCATAGACAAGGA 3325  
Qy 365 -----CCCTGAATTTCCCGCAGATCACTCTGTGGCAGCGCCCTTGGTAGCATCAAGG 418  
Db 3326 CTGTATCTTTAGCTTCCCTCAGATCACTCTTTTGGCAGCGACCCCTCTCGTCAATAAAGA 3385  
Qy 419 TGGCGCGCAGATCAAGAGGCGCTGTGGACACCGGCGCGAGACACCGCTGTCTGGAGG 478  
Db 3386 TAGGGGCGCAGCTGAAGAGGCGCTTCTAGACACCGGCGCGCAGACACCGCTGTCTGGAG 3445  
Qy 479 AGATGAGCTGCCCGCAAGTGGGAAGCCCAAGATGATCGGGCGCATCGCGGCTTTCATCA 538  
Db 3446 AGATGAACCTGCCCGCGCGCTGGAAGCCCAAGATGATCGGGCGCATCGGGGCTTTCATCA 3505  
Qy 539 AGGTGCGCGAGTAGCAGACAGATCTCGATCGAGATCTCGGGCAGAGAGCCATTCGGCACCG 598

Db 3506 AGTGGGCCAGTACGACACAGATCTGTATCGAGATCTCGGCCACAAAGGCCATCGGCACCG 3565  
QY 599 TGTGATCGGCCCCACCCCGTGAACATCATCGGCCCGCAACATGCTGACCCAGCTGGGCT 658  
Db 3566 TGTGTGGTGGGCCCGCACCCCGTGAACATCATCGGCCCGCAACCTGCTGACCCAGATCGGCT 3625  
QY 659 GCACCTCTGAATCTTCCCATCAGCCCATCAGAGCCGTGCGCGTGAAGCTGAAGCCCGGCA 718  
Db 3626 GCACCTCTGAATCTTCCCATCAGCCCATCAGAGCCGTGCGCGTGAAGCTGAAGCCCGGCA 3685  
QY 719 TGGACGGCCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAAGATCAAGGCCCTGACCG 778  
Db 3686 TGGACGGCCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAAGATCAAGGCCCTGCTGG 3745  
QY 779 CATCTGCGAGGATGGAAGAGGAGGCGCAAGATCAACAGATCGGCCCGCGAAGCCCT 838  
Db 3746 AGATCTGCAACCGAGATGGAAGAGGAGGCGCAAGATCAACAGATCGGCCCGCGAAGCCCT 3805  
QY 839 ACAACACCCCGCTGTTCCCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGGTGG 898  
Db 3806 ACAACACCCCGCTGTTCCCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGGTGG 3865  
QY 899 ACTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGCAAGTGGGCATCCCC 958  
Db 3866 ACTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGCAAGTGGGCATCCCC 3925  
QY 959 ACCCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGAGCGTGGGCGAGCCCTACT 1018  
Db 3926 ACCCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGAGCGTGGGCGAGCCCTACT 3985  
QY 1019 TCAGCGTGGCCCTGGACGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGATCA 1078  
Db 3986 TCAGCGTGGCCCTGGACGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGATCA 4045  
QY 1079 ACAACGAGACCCCGGATCCGCTACCAAGTACCAAGCTGTGCGCCCGAGGGTGAAGGGCA 1138  
Db 4046 ACAACGAGACCCCGGATCCGCTACCAAGTACCAAGCTGTGCGCCCGAGGGTGAAGGGCA 4105  
QY 1139 GCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCCCGGCCCGCAACC 1198  
Db 4106 GCCCGGCGCATCTTCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCCCGAAGCAAGC 4165  
QY 1199 CCAGAGTCTGATCTACCACTACATGACGACCTGTACGTGGCGAGCGACTCGAGATCG 1258  
Db 4166 CCAGATCTGATCTACCACTACATGACGACCTGTACGTGGCGAGCGACTCGAGATCG 4225  
QY 1259 GCAGCAACCGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCGCTGGGGTTCAACCA 1318  
Db 4226 GCAGCAACCGCGCAAGATCGAGAGCTGCGCGCAGCACTGTGCGCTGGGGTTCAACCA 4285  
QY 1319 CCCCCGACAAAGACCAAGAGAGGAGCCCTTCTGTGGATGGGCTAGAGCTGCACC 1378  
Db 4286 CCCCCGACAAAGAGACCAAGAGAGGAGCCCTTCTGTGGATGGGCTAGAGCTGCACC 4345  
QY 1379 CCAGCAAGTGAACCGGTGACCCCATCGAGCTGCGCGAGAGGAGAGCTGGAACCGTGAACG 1438  
Db 4346 CCAGCAAGTGAACCGGTGACCCCATCGAGCTGCGCGAGAGGAGAGCTGGAACCGTGAACG 4405  
QY 1439 ACATCCAGAGCTGGTGGGCAAGCTGAATCTGGGCCAGCCAGATCTACCCCGCATCAAGG 1498  
Db 4406 ACATCCAGAGCTGGTGGGCAAGCTGAATCTGGGCCAGCCAGATCTACCCCGCATCAAGG 4465  
QY 1499 TGGCCAGCTGTGAAGCTCTGCGGCGCGCGAGCCCTGACCGCATGCTGCCCTGA 1558  
Db 4466 TGGCCAGCTGTGAAGCTCTGCGGCGCGCGAGCCCTGACCGAGGTGGTGGCCCTGA 4525  
QY 1559 CCAGAGGCGCGAGCTGGAGCTGGCCGAGAACCGCAGATCTCTGCGGAGCCCGTGCACG 1618  
Db 4526 CCAGAGGCGCGAGCTGGAGCTGGCCGAGAACCGCAGATCTCTGAGAGAGCCCGTGCACG 4585  
QY 1619 GCGTGTACTACGACCCCGAGCAAGGACTGTGTGGCCGAGATCTCCAGAGCAGGGCCAGCAC 1678  
Db 4586 GCGTGTACTACGACCCCGAGCAAGGACTGTGTGGCCGAGATCTCCAGAGCAGGGCCAGGGCC 4645

QY 1679 AGTGGACCTACAGATCTACAGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCA 1738  
Db 4646 AGTGGACCTACAGATCTACAGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCC 4705  
QY 1739 AGATGCGCACCGCCCAACACGAGCTGAAGCAGCTGACCGAGGCGGTGACAGAAGATCG 1798  
Db 4706 GCATGAAGGGCGCCCAACACGAGCTGAAGCAGCTGACCGAGGCGGTGACAGAAGATCG 4765  
QY 1799 CATTGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCCGCTGCCCATCCAGNAG 1858  
Db 4766 CCACCGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTTCAAGCTGCCCATCCAGAGG 4825  
QY 1859 AGACCTGGGAGACCTGCTGGACCGACTACTTGGCAGCGCACTTGGATCCCGAGTGGGAGT 1918  
Db 4826 AGACCTGGGAGCGCTGCTGGACCGAGTACTGGGAGCGCACTTGGATCCCGAGTGGGAGT 4885  
QY 1919 TCGTGAACACCCCGCTGGTGAAGCTGTGTACCAAGTGGGAGAGGAGCCCATCATCG 1978  
Db 4886 TCGTGAACACCCCGCTGGTGAAGCTGTGTACCAAGTGGGAGAGGAGCCCATCATCG 4945  
QY 1979 GCGCCGAGACCTTCTAGTGGACGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038  
Db 4946 GCGCCGAGACCTTCTAGTGGACGGCGCGCCCAACCGCGAGACCAAGTGGGCAAGGCCG 5005  
QY 2039 GCTACGTGACCGACCGCGCGCGCAGAAAGATCTGTAGCCTTGACCGAGACCAACACAGA 2098  
Db 5006 GCTACGTGACCGACCGCGCGCGCAGAAAGTGTGTGCTTACCGACCAACCAACAGA 5065  
QY 2099 AGACCGAGCTGCAAGGCCATCTCAGCTGGCCCTGACGAGCAGCGCAGAGGTGAACATCG 2158  
Db 5066 AGACCGAGCTGCAAGGCCATCTCAGCTGGCCCTGACGAGCAGCGCCTGGAGGTGAACATCG 5125  
QY 2159 TGACCGAGCAGTACGCGCTGGGATCTTCCAGCCCGAGCCCGCGAGAGCGAGCGG 2218  
Db 5126 TGAACCGAGCAGTACGCGCTGGGATCTTCCAGCCCGAGCCCGCGAGAGCGAGCGG 5185  
QY 2219 AGCTGTGTGAACCAAGTATCTGAGCAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGG 2278  
Db 5186 AGCTGTGTGAACCAAGTATCTGAGCAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGG 5245  
QY 2279 TGCCCGCCCAAGAGGATCGCGGCAACGAGCAGATCGCAAGCTGGTGAAGAGGCA 2338  
Db 5246 TGCCCGCCCAAGAGGATCGCGGCAACGAGCAGCTGAGCGCCTGGTGAAGCGCGCA 5305  
QY 2339 TCCGCAAGTGTCTGTTCTTGGACGGCATCGATGGCGGCATCTGTATCTACAGTA 2393  
Db 5306 TCCGCAAGTGTCTGTTCTTGGACGGCATCGCAAGGCGCCAGGAGGAGCACAGAA 5360

## RESULT 11

AX427927 LOCUS 12411 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 165 from Patent WO0232943.  
ACCESSION AX427927  
VERSION AX427927.1 GI:21538014  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Huang, Y. and Nabel, G. J.  
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization  
JOURNAL Patent: WO 0232943-A 165 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
FEATURES  
source  
1. .12411  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN



Query Match			
Best Local Similarity 82.1%; Score 2027; DB 6; Length 12411;			
Matches 2194; Conservative 0; Mismatches 185; Indels 16; Gaps 4;			
Qy	14	TGGCCGAGGCGCATGAGCCAGG-CCACCAGGCGCAACATCTGATCGAGCGCAGCAACT	70
Db	2967	TGGCCGAGGCGCATGAGCGAGTGACCAACAGCGCCACCATCATGATGAGCGCGCAACT	3026
Qy	71	TCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGGCACATCGCCC	130
Db	3027	TCCGCAACCGAGCGCAAGATCGTGAAGTCTTCAACTGCGGCAAGAGGCGGCACACCGCCC	3086
Qy	131	GCAACTGCGCGCGCCCGCGGCAAGAGGCTGCTGGAAGTGGCGGCAAGAGGCGGCACAGGA	190
Db	3087	GCAACTGCGCGCGCCCGCGGCAAGAGGCTGCTGGAAGTGGCGGCAAGAGGCGGCACAGGA	3146
Qy	191	TGAAGGATGCAACGAGCGCCAGGCGCAACTTCTTCCGCGAGGACTGCGCCCTTCCCGCAGG	250
Db	3147	TGAAGGATGCAACGAGCGGACAGGCTAA-TTTTITTTAGGAAAGATCTGCGCCCTTCCCAAG	3205
Qy	251	GCAAGGCGCGGAGTCTCCCGAGCGAGCAGAAACCGGCCAACAGCCCAACAGCGCGCAGC	310
Db	3206	GGAAGGCGCAGGAATTTCTTTCAGAGCAGACCAAGCCCAACAGCCCAACAGAGAGAGC	3265
Qy	311	TGCAAGTGGCGG-----CGAACCCCGCGAGCGAGGCGCGCGCGAGCGCGCAGGGCA	364
Db	3266	TTCAAGTTTGGGGAAGAGACAACAACCTCCCTCTCAGAAAGCAGAGCGCGATAGACAAGGAA	3325
Qy	365	-----CCCTGAACCTCCCGCAGATCACTCTGTCGCGAGCGCGCCCTGGTGAAGATCAAGG	418
Db	3326	CTGTATCTTTAGTCTTCCCTCAGATCACTCTTTGGCAGCGACCTCTGTCACAATAAGA	3385
Qy	419	TGGCGCGCCAGATCAAGAGGCGCCCTGCTGGACACCGCGCGCGACGACACCGTGTGGAGG	478
Db	3386	TAGGGGCGCAGCTGAAAGAGGCGCTTCTAGACACCGGCGCGACGACACCGTGTGGAGG	3445
Qy	479	AGATGAGCCTGCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCAATCGGCGGCTTCAACA	538
Db	3446	AGATGAACCTGCCCGCGCGCTGGAAGCCCAAGATGATCGGCGGCAATCGGCGGCTTCAACA	3505
Qy	539	AGTGCGCCAGTACGACAGATCTGATCGAGATCTGGGCGCAAGAGCCATCGGACCG	598
Db	3506	AGTGCGCCAGTACGACAGATCTGATCGAGATCTGGGCGCAAGAGCCATCGGACCG	3565
Qy	599	TGCTGATCGCGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCT	658
Db	3566	TGCTGATCGCGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGATCGGCT	3625
Qy	659	GCACCTCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCA	718
Db	3626	GCACCTCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCA	3685
Qy	719	TGGAGCGCCCAAGTGAAGAGTGGCCCTTGAACGAGGAGAGATCAAGGCCCTGACCG	778
Db	3686	TGGAGCGCCCAAGTGAAGAGTGGCCCTTGAACGAGGAGAGATCAAGGCCCTTGTGG	3745
Qy	779	CCATCTGCGAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGCGCCCGGAGAACCCCT	838
Db	3746	AGATCTGCAACGAGATGGAAGAGGAGGCAAGATCAGCAAGATCGGCCCCCGAGAACCCCT	3805
Qy	839	ACAAACCCCGTGTTCGCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGGTG	898
Db	3806	ACAAACCCCGTGTTCGCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGGTG	3865
Qy	899	ACTTCCGCGAGCTGAACAGGCGACCCAGGACTTCTGGAGGTGACGTGGGATCCCC	958
Db	3866	ACTTCCGCGAGCTGAACAGGCGACCCAGGACTTCTGGAGGTGACGTGGGATCCCC	3925
Qy	959	ACCCCGCGGCTTGAAGAAAGAGAGCGGTGACCGTGTGGAGCTGGGCGAGCGCTACT	1018
Db	3926	ACCCCGCGGCTTGAAGAAAGAGAGCGGTGACCGTGTGGAGCTGGGCGAGCGCTACT	3985

Qy	1019	TCAGCGTGCCTCGGACGAGGACTTCCGCAAGTACACCGCCTTACCACTCCCGACATCA	1078
Db	3986	TCAGCGTGCCTCGGACGAGGACTTCCGCAAGTACACCGCCTTACCACTCCCGACATCA	4045
Qy	1079	ACAAAGAGAGAGCCCGGATCGCTACAGTACAAAGTCTGCTCCCGAGGCTTGAAGGCA	1138
Db	4046	ACAAAGAGAGAGCCCGGATCGCTACAGTACAAAGTCTGCTCCCGAGGCTTGAAGGCA	4105
Qy	1139	GCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACC	1198
Db	4106	GCCCGAGCATCTTCCAGTGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCAAGCAGAAC	4165
Qy	1199	CCGAGATCTGATCTACAGTACATGAGCAGCTGTAGCTGGGAGGAGCCTGGAGATCG	1258
Db	4166	CCGACATCTGATCTACAGTACATGAGCAGCTGTAGCTGGGAGGAGCCTGGAGATCG	4225
Qy	1259	GCCAGACCGCGCCAAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCTGGGCTTTCACCA	1318
Db	4226	GCCAGACCGCGCCAAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCTGGGCTTTCACCA	4285
Qy	1319	CCCCCGAACAAGACCAAGAGGAGCCCTTCTGTGTGATGGGCTTACGAGTGTGACC	1378
Db	4286	CCCCCGAACAAGACCAAGAGGAGCCCTTCTGTGTGATGGGCTTACGAGTGTGACC	4345
Qy	1379	CCGACAGTGCAGCTGCGGCCATCGAGCTGCGCGAGAGAGAGCTGGAGCCGTGAAG	1438
Db	4346	CCGACAGTGCAGCTGCGGCCATCGAGCTGCGCGAGAGAGAGCTGGAGCCGTGAAG	4405
Qy	1439	ACATCCAGAAGCTGCTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGCTCAAGG	1498
Db	4406	ACATCCAGAAGCTGCTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGCTCAAGG	4465
Qy	1499	TGCGCCAGCTGTGCAAGCTGCTGCGCGCGCCCAAGGCGCTGACCGACATCGTGCCTTGA	1558
Db	4466	TGCGCCAGCTGTGCAAGCTGCTGCGCGCGCCCAAGGCGCTGACCGAGCTGCTGCCTTGA	4525
Qy	1559	CCGAGAGGCGGAGCTGAGCTGCGCGAGAAACCGCGAGATCTCTGGCGAGGCGCGTGACG	1618
Db	4526	CCGAGAGGCGGAGCTGAGCTGCGCGAGAAACCGCGAGATCTCTGAAGAGGCGCGTGACG	4585
Qy	1619	GGGTGACTACGACCCCGAGCAAGGACCTGGTGGCGGAGATCCAGAGCAGGCGCCAGACC	1678
Db	4586	GGGTGACTACGACCCCGAGCAAGGACCTGGTGGCGGAGATCCAGAGCAGGCGCCAGACC	4645
Qy	1679	AGTGAACCTTACAGATCTACCGAGAGCCCTTCAAGAACCTTCAAGAACCTGAGACCGCC	1738
Db	4646	AGTGAACCTTACAGATCTACCGAGAGCCCTTCAAGAACCTTCAAGAACCTGAGACCGCC	4705
Qy	1739	AGATGCGCACCGGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGAGAGATCG	1798
Db	4706	GCATGAAGGCGCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGAGAGATCG	4765
Qy	1799	CCATGAGAGCATCGTGTATCTGGGCGAAGACCCCAAGTTCGCTGCCATCCAGAGG	1858
Db	4766	CCATGAGAGCATCGTGTATCTGGGCGAAGACCCCAAGTTCGCTGCCATCCAGAGG	4825
Qy	1859	AGACTGGGAGACCTGCTGAGACCGACTACTTGGCAGGCGCCCTGGATCCCGAGTGGGAGT	1918
Db	4826	AGACTGGGAGACCTGCTGAGACCGACTACTTGGCAGGCGCCCTGGATCCCGAGTGGGAGT	4885
Qy	1919	TCGTGAACACCCCGCTGCTGAGCTGTGGTACAGCTGGAGAGAGAGGCGCCATCATCG	1978
Db	4886	TCGTGAACACCCCGCTGCTGAGCTGTGGTACAGCTGGAGAGAGAGGCGCCATCATCG	4945
Qy	1979	GCGCGGAGACCTTCTAGCTGAGCGCGCCGCAACCGGAGAGACCAAGATCGGAGGCGCG	2038
Db	4946	GCGCGGAGACCTTCTAGCTGAGCGCGCCGCAACCGGAGAGACCAAGATCGGAGGCGCG	5005
Qy	2039	GCTACGTGAACCGAGCGGCGCGGAGAGATCTGTGAGCTGACCGAGACCAACCAAGAG	2098
Db	5006	GCTACGTGAACCGAGCGGCGCGGAGAGATCTGTGAGCTGACCGAGACCAACCAAGAG	5065
Qy	2099	AGACCGAGCTGACGCGCATCCAGCTGCGCTTGCAGGACAGCGGAGCGAGGTTGAACATCG	2158



Db	5066	AGACCGAGCTGCAGGCATCCACCTGGCCCTGCAGGACGCGCCTGGAGGTGAACATCG	5125	
Qy	2159	TGACCGACAGCCAGTAGTACGCCCTGGGCACTCATCCAGGCCGAGCCGACAGAGCGAGCG	2218	
Db	5126	TGACCGACAGCCAGTAGTACGCCCTGGGCACTCATCCAGGCCGAGCCGACAGAGCGAGCG	5185	
Qy	2219	AGCTGGTGAACCAAGATCATCGACGAGTATCAAGAGGAGGAGAGGTGTACTCGAGCTGGG	2278	
Db	5186	AGCTGGTGAACCAAGATCATCGACGAGTATCAAGAGGAGGAGAGGTGTACTCGAGCTGGG	5245	
Qy	2279	TGCCCCGCCCAAGAGGCGATCGCGGCAACGAGCAGATCGCAAGCTGGTGAGCAAGGGCA	2338	
Db	5246	TGCCCCGCCCAAGAGGCGATCGCGGCAACGAGCAGTGGAGCGCGCTGGTGAGCGCGCA	5305	
Qy	2339	TCCGCAAGGTGCTTCTCGGACGGGATCGATCGCGGCGATCGTGATCTACCACTA	2393	
Db	5306	TCCGCAAGGTGCTTCTCGGACGGGATCGATCGCGGCGATCGTGATCTACCACTA	5360	
RESULT 12				
AX427926				
LOCUS				
DEFINITION				
AX427926				
ACCESSION				
VERSION				
AX427926.1				
GI:21538013				
KEYWORDS				
synthetic construct				
SOURCE				
ORGANISM				
synthetic construct				
other sequences; artificial sequences.				
REFERENCE				
1				
Huang Y. and Nabel G.J.				
Modifications of hiv env, gag, and pol enhance immunogenicity for				
Genetic immunization				
Patent: WO 0232943-A 164 25-APR-2002;				
GOVERNMENT OF THE UNITED STATES (US)				
LOCATION/Qualifiers				
1. -9194				
/organism="synthetic construct"				
/mol_type="unassigned DNA"				
/db_xref="taxon:32630"				
/notes="plasmid pVR1012x/s containing HIV genes"				
FEATURES				
source				
ORIGIN				
Query Match				
Best Local Similarity 82.0%; Score 2025.4; DB 6; Length 9194;				
Matches 2193; Conservative 0; Mismatches 186; Indels 16; Gaps 4;				
Qy	14	TGCCCCGAGGCGCATGAGCGAGG---CCACACAGCGCCAACTCTCTGATGCGAGCGCAGCAACT	70	
Db	2967	TGCCCCGAGGCGCATGAGCGAGGCGACCAAGCGCCACCATCATGATGCGAGCGCGCACT	3026	
Qy	71	TCAAGGGCCCCAAGCGCGATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATCGCCC	130	
Db	3027	TCCGCAACACAGCGCAAGATCGTGAAGTGTCTTCAACTGCGGCAAGGAGGGCCACACCGCCC	3086	
Qy	131	GCAACTGCGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTCCGCGAGGAGGGCCACAGGA	190	
Db	3087	GCAACTGCGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTCCGCGAGGAGGGCCACAGGA	3146	
Qy	191	TGAAGGACTGCACTGAGCGCGAGCGCAACTTCTTCCGCGAGGAGGAGGAGGAGGAGGAGG	250	
Db	3147	TGAAGGACTGCACTGAGCGCGAGCGCAACTTCTTCCGCGAGGAGGAGGAGGAGGAGGAGG	3205	
Qy	251	GCAAGGCGCGCGAGTTCCTCCAGCGAGCAGCAACCGCGCCAAAGCGCCGAGCGCGAGCG	310	
Db	3206	GGAAGGCGCGCGAGTTCCTCCAGCGAGCAGCAACCGCGCCAAAGCGCCGAGCGCGAGCG	3265	
Qy	311	TGCAGGTGCGCGG-----CGACAAACCGCGAGCGAGGCGCGCGAGCGCGAGGGCA	364	
Db	3266	TTCAGGTGCGCGGAAAGAGCAACAACTCTCTCTCAGAAAGCAGGAGCGGATAGACAAAGAA	3325	
Qy	365	-----CCCTGAACTTCCCCCCAGATCACCCCTGTGGCAGCGCCCTTGGTGAGCATCAAGG	418	

Db	3326	CTGTATCTCTTTAGCTTTCCTCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAACAATAAGA	3385	
Qy	419	TGGGCGCGCAGATCAAGAGGAGGCGCTCTGTGACACACCGGCGCGACACCGCTGCTGGAGG	478	
Db	3386	TAGGGGGGCGAGCTGAAAGGAGGCGCTTCTAGACACACCGGCGCGACACCGCTGCTGGAGG	3445	
Qy	479	AGATGAGCCTTGCCCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCA	538	
Db	3446	AGATGAACCTTGCCCGGCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCA	3505	
Qy	539	AGTGTGCGCCAGTAGTACGACAGATCTCTGATCGAGATCTTGCGGCAAGAGGCGCATCGGACCG	598	
Db	3506	AGTGTGCGCCAGTAGTACGACAGATCTCTGATCGAGATCTTGCGGCAAGAGGCGCATCGGACCG	3565	
Qy	599	TGCTGATCGGCGCCACCCCGGTGAACATCATCGGCGGCACATCATGCTGACCGCAGCTGGGCT	658	
Db	3566	TGCTGATCGGCGCCACCCCGGTGAACATCATCGGCGGCACATCATGCTGACCGCAGCTGGGCT	3625	
Qy	659	GCACCTGGAACCTTCCCGCATCAGCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGGCA	718	
Db	3626	GCACCTGGAACCTTCCCGCATCAGCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGGCA	3685	
Qy	719	TGACCGGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCCCTTGACCG	778	
Db	3686	TGGACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCCCTTGGTGG	3745	
Qy	779	CMATCTGCGAGGAGATGGAGAGGAGGCGGCAAGATCAACCAAGATCGGCGCCCGAGAACCCCT	838	
Db	3746	AGATCTGCAACCGAGATGGAGAGGAGGCGGCAAGATCAAGCAAGATCGGCGCCCGAGAACCCCT	3805	
Qy	839	ACAACACCCCGCTGTTTCGCCATCAAGAAGAGGACAGACCAAGTGGCGCAAGCTGGTGG	898	
Db	3806	ACAACACCCCGCTGTTTCGCCATCAAGAAGAGGACAGACCAAGTGGCGCAAGCTGGTGG	3865	
Qy	899	ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCGAGCTGGGCAATCCCC	958	
Db	3866	ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCGAGCTGGGCAATCCCC	3925	
Qy	959	ACCCCGCGCGCTGAAGAGAGGAGCGGTGACCGTGTGCGAGCTGGGCGAGCGCTACT	1018	
Db	3926	ACCCCGCGCGCTGAAGAGAGGAGCGGTGACCGTGTGCGAGCTGGGCGAGCGCTACT	3985	
Qy	1019	TCAGCTGCGCCCTGGACGAGGACTTCCGCAAGTACACACCGCTTTCACCATCCCCAGCATCA	1078	
Db	3986	TCAGCTGCGCCCTGGACGAGGACTTTCGCAAGTACACACCGCTTTCACCATCCCCAGCATCA	4045	
Qy	1079	ACAACGAGACCCCGCGCATCCGCTACAGTACAAACGTTGCTGCCCGAGGGCTGGAAGGGCA	1138	
Db	4046	ACAACGAGACCCCGCGCATCCGCTACAGTACAAACGTTGCTGCCCGAGGGCTGGAAGGGCA	4105	
Qy	1139	GCCCCAGCATCTTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACC	1198	
Db	4106	GCCCCAGCATCTTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACC	4165	
Qy	1199	CCGAGATCTGTACTACCATGAGACGACCTGTACGTTGGGCGAGCGACCTGGAGATCG	1258	
Db	4166	CCGATCTGTGATCTTACCATGATACGACCACTGTACGTTGGGCGAGCGACCTGGAGATCG	4225	
Qy	1259	GCCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTGCTGCGCTGGGGCTTTCACCA	1318	
Db	4226	GCCAGCACCGCGCAAGATCGAGGAGCTGCGCCAGCACTGCTGCGCTGGGGCTTTCACCA	4285	
Qy	1319	CCCCCGCAAGAGCACCAAGAGGAGGCGCCCTTCTGTGGATGGGCTAGAGCTGCACC	1378	
Db	4286	CCCCCGCAAGAGCACCAAGAGGAGGCGCCCTTCTGTGGATGGGCTAGAGCTGCACC	4345	
Qy	1379	CCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGAGTGCACCGTGAACG	1438	
Db	4346	CCGACAAAGTGGACCGTGCAGCCCATCGTGTGCTGCCGAGAGAGACAGCTGGACCGTGAACG	4405	
Qy	1439	ACATCCAGAGCTGGTGGGCAAGCTGAACTTGGGCGCAGCCAGATCTTACCCCGCGCATCAAGG	1498	

Db 4406 ACATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGCAGGCAGATCTACGCCGGCATCAAGG 4465  
Qy 1499 TGCGCCAGCTGTCAAGCTGTGCGCGCGCCAAAGGCCCTCAGCAGCATCGTGCCCTGA 1558  
Db 4466 TGCGCCAGCTGTCAAGCTGTGCGCGCCAAAGGCCCTCAGCAGCATCGTGCCCTGA 4525  
Qy 1559 CCGAGGAGGCCAGCTGGAGTGGCCGAGAACCCGGAGATCTCTGGCGAGGCCCTGACG 1618  
Db 4526 CCGAGGAGGCCAGCTGGAGTGGCCGAGAACCCGGAGATCTCTGAAGAGGCCCTGACG 4585  
Qy 1619 GCGTGTACTACGACCCAGCAAGCACTGGTGGCGCGAGATCCAGAAGCAGGGCCACAGCC 1678  
Db 4586 GCGTGTACTACGACCCAGCAAGCACTGTGCGCGAGATCCAGAAGCAGGGCCAGGGCC 4645  
Qy 1679 AGTGGACTTACAGATCTACAGAGGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738  
Db 4646 AGTGGACTTACAGATCTTACCAGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCC 4705  
Qy 1739 AGATGGCAGCGCCACACACAGCAAGCTGAAGCAGCTGACCGAGGCCGTGCAGAGATCG 1798  
Db 4706 GCATGAAGGGCGCCACACCAACAGCAGTGAAGCAGCTGACCGAGGCCGTGCAGAGATCG 4765  
Qy 1799 CCATGGAGAGCATCGTGTATCTGGGGCAAGACCCCAAGTTCGGCTCGCCATCCAGAGG 1858  
Db 4766 CCACCGAGAGCATCGTGTATCTGGGGCAAGACCCCAAGTTCAGCTGCCATCCAGAGG 4825  
Qy 1859 AGACTGGGAGACCTGGTGAACCGACTACTGGCAGGCCACCTGTGATCCCGAGTGGGAGT 1918  
Db 4826 AGACTGGGAGCCCTGGTGGACCGAGTACTGGCAGGCCACCTGTGATCCCGAGTGGGAGT 4885  
Qy 1919 TCGTGAACACCCCGCTGGTGAAGCTGTGGTACAGCTGGAGAGAGGCCATCATCG 1978  
Db 4886 TCGTGAACACCCCGCTGGTGAAGCTGTGGTACAGCTGGAGAGAGGCCATCATCG 4945  
Qy 1979 GCGCCGAGACCTTCTAGCTGAACGCGCGCCAGCAAGCTGACCAAGATCGGCAAGGCCG 2038  
Db 4946 GCGCCGAGACCTTCTAGCTGAACGCGCGCCAGCAAGCTGACCAAGATCGGCAAGGCCG 5005  
Qy 2039 GCTACGTGACCGACCGGGCGCGCAGAGAGTCTGTAGCTGACCGAGACCAAGATCGGCAAG 2098  
Db 5006 GCTACGTGACCGACCGGGCGCGCAGAGAGTCTGTAGCTGACCGAGACCAAGATCGGCAAG 5065  
Qy 2099 AGACCGAGCTGACGGCCATTCAGCTGGCCCTGCAAGGACAGCGGAGCGAGTGAACATCG 2158  
Db 5066 AGACCGAGCTGACGGCCATTCACCTGGCCCTGCAAGGACAGCGGCGCTGGAGGTGAACATCG 5125  
Qy 2159 TGACCGACAGCCAGTACGCTGGGCGCATCTCCAGGCGCCAGCCGACAGAGCGAGAGCG 2218  
Db 5126 TGACCGACAGCCAGTACGCTGGGCGCATCTCCAGGCGCCAGCCGACAGAGCGAGAGCG 5185  
Qy 2219 AGCTGGTGAACCCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGG 2278  
Db 5186 AGCTGGTGAACCCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGG 5245  
Qy 2279 TGCCCGCCCAAGAGGATCGCGGCAAGCAGCAGATCGCAAGCAGCTGGTGAAGCAAGGGCA 2338  
Db 5246 TGCCCGCCCAAGAGGATCGCGGCAAGCAGCAGGTGGACGGCTGGTGAAGCGCCGCA 5305  
Qy 2339 TCCGCAAGGTGCTGTTCTTGTGACGGCATCGATCGCGGCATCGTATCTACAGTA 2393  
Db 5306 TCCGCAAGGTGCTGTTCTTGTGACGGCATCGCAAGAGGCCAGAGGAGGACAGAGAA 5360

RESULT 13  
LOCUS BD263704  
DEFINITION Improved expression of HIV polypeptides and production of  
virus-like particles.  
ACCESSION BD263704  
VERSION BD263704.1 GI:33073472  
KEYWORDS JP 2002533124-A/71.  
SOURCE synthetic construct  
ORGANISM

other sequences; artificial sequences.  
1 (bases 1 to 2306)  
Barnett, S., Megede, J.Z., Srivastava, I., Lian, Y., Hartog, K., Liu, H.,  
Greer, C., Selby, M. and Walker, C.  
Improved expression of HIV polypeptides and production of  
virus-like particles  
Patent: JP 2002533124-A 71 08-OCT-2002;  
CHIRON CORP  
OS Artificial Sequence  
PN JP 2002533124-A/71  
PD 08-OCT-2002  
PF 30-DEC-1999 JP 2000591193  
PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/158471 PI  
SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI  
KARIN HARTOG,  
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC  
CI2N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC  
CI2N5/10,  
PC CI2N7/00, CI2P21/02, CI2N15/00, CI2N5/00, A61K37/02 CC  
Description of Artificial Sequence: FS(-). protmod.RTopt.YM PH Key  
Location/Qualifiers  
FT source 1..2306  
FT /organism='Artificial Sequence'.  
FEATURES  
source Location/Qualifiers  
1..2306  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'  
ORIGIN  
Query Match 82.0%; Score 2025.2; DB 6; Length 2306;  
Best Local Similarity 93.3%; Pred. No. 6.1e-203;  
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
Qy 170 GCGGCAAGGAGGCGCCACAGATGAAGGACTGACCGAGCGCGCAGCAACTTCTTCGCGC 229  
Db 1 GCGGCGCGGAGGACCAATGAAGATTGCATGAGACAGAGCTAATTTCTTCGCGC 60  
Qy 230 AGGACCTTGGCTTCCCCAGGCGAAGCGCGGAGTTCCCCAGGAGCAGAACCGCGCA 289  
Db 61 AGGACCTTGGCTTCTCTGCGGCGAAGCGCGGAGTTTCAGCAGCGAGCAGACCGCGCA 120  
Qy 290 ACAGCCCCACAGCGCGGAGCTGAGTGGCGGCGG-----ACAAACCGCGAGCGAGG 343  
Db 121 ACAGCCCCACCGCGCGAGCTGCGAGTGTGGGCGCGCGAGAAACAGCGCTGAGCGAGG 180  
Qy 344 CCGGCGCGGAGCGCGGCGGCGCGCTG-----ACTTCCCCAGATCACCTGTGGGCGC 397  
Db 181 CCGGCGCGGAGCGCGGCGGCGCGTGTGAGCTTCAACTTCCCCCAGATCACCTGTGGGCGC 240  
Qy 398 GCCCCTGTGTGAGCATCAAGGTGGGCGGCGCAGATCAAGGAGGCCCTGTGTGACACCGCGC 457  
Db 241 GCCCCTGTGTGACCATCAGGATCGGCGCGCAGCTCAAGGAGGCGCTGCTGACACCGCGC 300  
Qy 458 CCGACGACACCGTGTGTGAGAGATGAGCTGCGCGGCGAAGTGAAGCCCAAGATGATCG 517  
Db 301 CCGACGACACCGTGTGTGAGAGATGAACCTGCGCGGCGAAGTGAAGCCCAAGATGATCG 360  
Qy 518 GCGGCATCGGCGGCTTCATCAAGGTGGCGCGCAGTACGACCGAGATCCTGATCGAGATCTGCG 577  
Db 361 GCGGAGTGGGGGCTTCAAGGTGGCGGAGTACGACCGAGATCCCCGTGGAGATCTGCG 420  
Qy 578 GCAAGAGGCGCATCGGACCGTGTGTATCGGCGCGCGCGCGCGTGAACATCATCGCGCGCA 637  
Db 421 GCCAAGGCGCATCGGACCGTGTGTGGGCGCGCGCGCGTGAACATCATCGCGCGCA 480  
Qy 638 ACATGTGACCCAGCTGGGCTGCACCTGTGAATTTCCCATCAGCCCCATCGAGACCGTGC 697  
Db 481 ACCTGTGACCCAGATCGGCTGCACCTGAACTTCCCATCAGCCCCATCGAGACCGGTGC 540  
Qy 698 CCGTGAAGCTGAAGCCCGGCATGAGCGGCGCGCGAGGTGAAGCGAGTGGCCCTCGACCGAGG 757  
Db 541 CCGTGAAGCTGAAGCCCGGCATGAGCGGCGCGCGCGCGCGTGAAGGTCAAGGTGGCCCTCGACCGAGG 600

QY 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAGAGAGGGCAAGATCACCA 817  
 Db 601 AGAAGATCAAGGCGCTGCTGGAGATCTGCACGAGATGGAGAGAGGGCAAGATCAGCA 660  
 QY 818 AGATCGGCGCCGAGAACCCCTCAACAACCCCGCTGTTTCGCCATCAAGAAGAAGACAGCA 877  
 Db 661 AGATCGGCGCCGAGAACCCCTCAACAACCCCGCTGTTTCGCCATCAAGAAGAAGACAGCA 720  
 QY 878 CCAAGTGGCGCAAGCTGGTGGATCTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 937  
 Db 721 CCAAGTGGCGCAAGCTGGTGGATCTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 780  
 QY 938 AGGTGAGCTGGGCGATCCCGCCACCCCGCGCGCTGAAAGAAAGAGAGGCTGACCGTGC 997  
 Db 781 AGGTGAGCTGGGCGATCCCGCCACCCCGCGCGCTGAAAGAAAGAGAGGCTGACCGTGC 840  
 QY 998 TGGACGTGGGCGAGCGCTACTTCAGCGTGCCTTCGACGAGGACTTTCGCAAGTACACCG 1057  
 Db 841 TGGACGTGGGCGAGCGCTACTTCAGCGTGCCTTCGACGAGGACTTTCGCAAGTACACCG 900  
 QY 1058 CTTTACCATCCGAGATCAACAACGAGACCCCGCGCATCCGCTACCAAGTACAACTGTC 1117  
 Db 901 CTTTACCATCCGAGATCAACAACGAGACCCCGCGCATCCGCTACCAAGTACAACTGTC 960  
 QY 1118 TGCCCGAGGGCTGGAAGGGGAGCCCGAGCATCTTCAGAGCAGCATGACCAAGATCCTGG 1177  
 Db 961 TGCCCGAGGGCTGGAAGGGGAGCCCGCGCATCTTCAGAGCAGCATGACCAAGATCCTGG 1020  
 QY 1178 AGCCCTTCGCGCGCGCAACCCCGAGATCGTGATCTACCAAGTACATGGACGACCTGTACG 1237  
 Db 1021 AGCCCTTCGCGCGCGCAACCCCGAGATCGTGATCTACCAAGTACATGGACGACCTGTACG 1074  
 QY 1238 TGGGAGCGACTGGAGATCGGCGAGCACCGCGCGCAAGATCGAGAGCTCGCGCAAGCAC 1297  
 Db 1075 TGGGAGCGACTGGAGATCGGCGAGCACCGCGCGCAAGATCGAGAGCTCGCGCAAGCAC 1134  
 QY 1298 TGCTGGCTGGGCTTCACCAACCCCGACAAGAGCACCAAGAAGAGGAGCCCGCTTCCTGT 1357  
 Db 1135 TGCTGGCTGGGCTTCACCAACCCCGACAAGAGCACCAAGAAGAGGAGCCCGCTTCCTGT 1194  
 QY 1358 GGATGGGCTACGAGCTGCACCCCGACAAAGTGGACCGGTGACGCCCATCGAGCTGCCGAGA 1417  
 Db 1195 GGATGGGCTACGAGCTGCACCCCGACAAAGTGGACCGGTGACGCCCATCGAGCTGCCGAGA 1254  
 QY 1418 AGAGAGCTGGAACGCTGGAACGATCCAGAGCTGGTGGGCAAGCTGGAAGTGGGCGAGCC 1477  
 Db 1255 AGGACAGCTGGACCGTGAACGATCCAGAGCTGGTGGGCAAGCTGGAAGTGGGCGAGCC 1314  
 QY 1478 AGATCTACCCCGGCGATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGGGCGCAAGGCC 1537  
 Db 1315 AGATCTACCCCGGCGATCAAGGTGAGAGCTGTGCAAGCTGTGCGGGCGCAAGGCC 1374  
 QY 1538 TGACCGAGCATCGTGCCCTGACCGAGGAGCGAGCTGGAGCTGGCGGAGAACCGCGAGA 1597  
 Db 1375 TGACCGAGGTGATCCCTCTGACCGAGGAGCGAGCTGGAGCTGGCGGAGAACCGCGAGA 1434  
 QY 1598 TCTGCGGAGCGGCTGACCGGCTGTACTAGACCCCGAGAGGAGCTGGTGGCGGAGA 1657  
 Db 1435 TCTGGAAGGAGCGGCTGACGAGGTGTACTAGACCCCGAGAGGAGCTGGTGGCGGAGA 1494  
 QY 1658 TCCAGAGCAGGCGCACGACGAGTGGACCTTACAGATCTACAGGAGCCCTTCAAGAAC 1717  
 Db 1495 TCCAGAGCAGGCGCACGAGGCGAGTGGACCTTACAGATCTACAGGAGCCCTTCAAGAAC 1554  
 QY 1718 TGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACACCAACGAGCTGAAGCAGCTGA 1777  
 Db 1555 TGAAGACCGGCAAGTACGCGCGATGCGCGCGGCCACACCAACGAGCTGAAGCAGCTGA 1614  
 QY 1778 CCGAGGCGTGCAGAGATCGCCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGT 1837  
 Db 1615 CCGAGGCGTGCAGAGAGTGAGCACCGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGT 1674

QY 1838 TCCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1897  
 Db 1675 TCAAGCTGCCCATCCAGAGGAGACCTGGGAGGAGCTGGTGGATGGAGTACTGGCAGGCCA 1734  
 QY 1898 CTTGGATTCCTCGAGTGGGAGTTCTGTGAACACACCCCGCTGGTGAAGCTGTGTACGAGC 1957  
 Db 1735 CTTGGATTCCTCGAGTGGGAGTTCTGTGAACACACCCCGCTGGTGAAGCTGTGTGTACGAGC 1794  
 QY 1958 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCG 2017  
 Db 1795 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCG 1854  
 QY 2018 AGACCAAGATTCGGCAAGGCGGCTAGCTGACCGACCGGGCGCGGAGAGATCGTAGGCC 2077  
 Db 1855 AGACCAAGCTGGGCAAGGCGGCTAGCTGACCGACCGGGCGCGGAGAGGTTGTGTAGCA 1914  
 QY 2078 TGAACGAGACCAACCAAGAGACGAGCTGACGAGCATCCAGCTGGGCGCTGGCCCTGAGGACA 2137  
 Db 1915 TCGCGGACACCAACCAAGAGACGAGCTGACGAGCATCCACCTGGGCGCTGAGGACA 1974  
 QY 2138 GCGGCGAGGAGTGAACATCGTACCGACGACGAGCTAGCGCTGGGCGATCATCCAGGCC 2197  
 Db 1975 GCGGCGTGGAGTGAACATCGTACCGACGACGAGCTAGCGCTGGGCGATCATCCAGGCC 2034  
 QY 2198 AGCCGACAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAG 2257  
 Db 2035 AGCCGACAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAG 2094  
 QY 2258 AGAAGTGTACTGAGCTGGTGCCTCCGCGCCCAAGGCGATCGGCGGCAACGAGCAGATCG 2317  
 Db 2095 AGAAGTGTACTGAGCTGGTGCCTCCGCGCCCAAGGCGATCGGCGGCAACGAGCAGGTGG 2154  
 QY 2318 ACAAGCTGTGTGAGCAAGGCGATCGCAAGTGTCTTCTTGGACGCGCATCGATGGCGGCA 2377  
 Db 2155 ACAAGCTGTGTGAGCGCGGATCCGCAAGTGTCTTCTTGAACGCGCATCGATGGCGGCA 2214  
 QY 2378 TCGTGATCTACCAAGTACATGAGCAGCTGTACGTGGGCGAGCGCGCTTAGGATCGATT 2437  
 Db 2215 TCGTGATCTACCAAGTACATGAGCAGCTGTACGTGGGCGAGCGCGCTTAGGATCGATT 2274  
 QY 2438 AAAAGCTTCCGGGGCTAGCACCGGTGAATTC 2469  
 Db 2275 AAAAGCTTCCGGGGCTAGCACCGGTGAATTC 2306

RESULT 14  
 CQ870574  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source

CQ870574 2306 bp DNA linear PAT 13-SEP-2004  
 Sequence 82 from Patent EP1433851.  
 CQ870574  
 CQ870574.1 GI:52000090  
 .  
 synthetic construct  
 synthetic construct  
 other sequences; artificial sequences.  
 1  
 Barnett, S., Zurmege, J., Srivastava, I., Lian, Y., Hartog, K.,  
 Liu, H., Greer, C., Selby, M. and Walker, C.  
 Improved expression of HIV polypeptides and production of  
 virus-like particles  
 Patent: EP 1433851-A 82 30-JUN-2004;  
 CHIRON CORPORATION (US)  
 Location/Qualifiers  
 1..2306  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="Description of Artificial Sequence:  
 FS(-).protmod.RTopt.YM"

ORIGIN  
 Query Match 82.0%; Score 2025.2; DB 6; Length 2306;  
 Best Local Similarity 93.3%; Pred. No. 6.1e-203;

		Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;			
Qy	170	CGCGCAAGGAGGCGCCACAGATGAAGAGTGCACCGAGCGCAGGCCAACTTCTTCGCG	229		
Db	1	CGCGCGCGGAGGACACCAATGAAGATTGCATGAGAGACAGGTAATTTCTTCGCG	60		
Qy	230	AGGACCTGGCTTCCCGCAGGCGAAGGCCCGGAGTTCCCGCAGGAGCAGAACCGCGCA	289		
Db	61	AGGACCTGGCTTCTTGAGGCGCAAGGCCCGGAGTTTCAGAGCGAGCAGACCCGCGCA	120		
Qy	290	ACAGCCCCCAGCGCGAGCTGCAGGTGCGCGCG-----ACAACCCCGCAGCGAGG	343		
Db	121	ACAGCCCCCAGCGCGAGCTGCAGGTGCGCGCGCGGCGGAGAAACAAGCTGAGCGAGG	180		
Qy	344	CGCGCGCGGAGCGCGAGGCGACCTG-----AACTTCCCGCAGATACCTGTGGCAGC	397		
Db	181	CGCGCGCGGAGCGCGAGGCGACCTGAGCTTCAACTTCCCGCAGATACCTGTGGCAGC	240		
Qy	398	GCCCCCTGGTGCAGATCAAGGTGGCGCGCAGATCAAGGAGGCCCTGCTGAGCACCGCG	457		
Db	241	GCCCCCTGGTGCAGATCAAGGTGGCGCGCAGCTCAAGGAGGCGCTGCTGAGCACCGCG	300		
Qy	458	CCGACGACACCGTGTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCG	517		
Db	301	CCGACGACACCGTGTGAGGAGATGAACCTGCGCGGAGTGAAGCCCAAGATGATCG	360		
Qy	518	CGCGCATCGCGGCTTCAACAAGGTGGCGCAGTACGACAGATCCTGATCGAGATCTGCG	577		
Db	361	CGCGCATCGCGGCTTCAATCAAGGTGGCGCAGTACGACAGATCCTGATCGAGATCTGCG	420		
Qy	578	GCAAGAGGCCATCGGCGACCGTGTGATCGGCCCGCACCCCGTGAACATCATCGGCGCA	637		
Db	421	GCCACAAGGCCATCGGCGACCGTGTGATCGGCCCGCACCCCGTGAACATCATCGGCGCA	480		
Qy	638	ACATGCTGACCCAGCTGGGCTGCACTTGAATTTCCCATCAGCCCATCGAGCCGTGC	697		
Db	481	ACCTGCTGACCCAGATGGCTGCACTTGAATTTCCCATCAGCCCATCGAGCCGTGC	540		
Qy	698	CGGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGGTGAAGCTGGGCCCTGACCGAGG	757		
Db	541	CGGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGGTGAAGCTGGGCCCTGACCGAGG	600		
Qy	758	AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAGAGGCGCAAGATCAACA	817		
Db	601	AGAAGATCAAGGCCCTGATGAGATCTGCAACCGAGATGAGAGAGGCGCAAGATCAACA	660		
Qy	818	AGATCGGCCCGAGAACCCCTACAAACCCCGGTTCGCCATCAAGAAGAAGCAGCA	877		
Db	661	AGATCGGCCCGAGAACCCCTACAAACCCCGGTTCGCCATCAAGAAGAAGCAGCA	720		
Qy	878	CCAAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG	937		
Db	721	CCAAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG	780		
Qy	938	AGGTGAGCTGGGATATCCCGACCCCGCGGCTGAAAGAGAGAGAGCGTGACCGTGC	997		
Db	781	AGGTGAGCTGGGATATCCCGACCCCGCGGCTGAAAGAGAGAGAGCGTGACCGTGC	840		
Qy	998	TGGACGTGGGCGACGCTTACTTACAGCTGCGCCCTGGACGAGGACTTCGCAAGTACACCG	1057		
Db	841	TGGACGTGGGCGACGCTTACTTACAGCTGCGCCCTGGACGAGGACTTCGCAAGTACACCG	900		
Qy	1058	CCTTCAACATCCCGCAGATCAACAACGAGACCCCGGCGATCCGCTACAGTACAACTGC	1117		
Db	901	CCTTCAACATCCCGCAGATCAACAACGAGACCCCGGCGATCCGCTACAGTACAACTGC	960		
Qy	1118	TGCCCCAGGGGTGAAGGGCGACCCCGACATCTTTCAGAGCAGATGATCAAGAATCTTGG	1177		
Db	961	TGCCCCAGGGGTGAAGGGCGACCCCGCATCTTTCAGAGCAGATGATCAAGAATCTTGG	1020		
Qy	1178	AGCCCTTCCGCGCGGCGACCCCGAGATCGTGATCTACAGTACATGGAACGACTGTACG	1237		
Db	1021	AGCCCTTCCGCGCGGCGACCCCGACATCGTGATCTACCA-----GGCCCCCTGTACG	1074		

Qy	1238	TGGCAGCGACCTCGAGATCGGCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCAC	1297		
Db	1075	TGGCAGCGACCTCGAGATCGGCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCAC	1134		
Qy	1298	TGCTGCCCTGGGCTTCAACACCCCGCAGAGAGCACCAAGAGAGCCCTTCTCTGT	1357		
Db	1135	TGCTGCCCTGGGCTTCAACACCCCGCAGAGAGCACCAAGAGAGCCCTTCTCTGT	1194		
Qy	1358	GGATGGGCTACGAGCTGACCCCGACCAAGTGGAGCCGTGACCCCATCGAGCTGCCGAGA	1417		
Db	1195	GGATGGGCTACGAGCTGACCCCGACCAAGTGGAGCCGTGACCCCATCATGTGCCGAGA	1254		
Qy	1418	AGGAGAGTGAACCGTGAAACGACATCCAGAAAGCTGTGGGCAAGCTGGAATCTGGGCGAGCC	1477		
Db	1255	AGGAGAGTGAACCGTGAAACGACATCCAGAAAGCTGTGGGCAAGCTGGAATCTGGGCGAGCC	1314		
Qy	1478	AGATCTACCCCGGCATCAAGGTGCGCAGCTGTGCAAGCTGCTGCGGCGCCCAAGGCC	1537		
Db	1315	AGATCTACCCCGGCATCAAGGTGAGCAGCTGTGCAAGCTGCTGCGGCGCCCAAGGCC	1374		
Qy	1538	TGACCGACATCTGTCCTCCCTGACCCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCGAGA	1597		
Db	1375	TGACCGAGGTGATCCCTCCCTGACCCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCGAGA	1434		
Qy	1598	TCCTGCGGAGCCCGTGACCGGCTGTACTACGACCCCGACCAAGGACTGTGTGGCGGAGA	1657		
Db	1435	TCCTGGAAGGAGCCCGTGACCGAGGTGTACTACGACCCCGACCAAGGACTGTGTGGCGGAGA	1494		
Qy	1658	TCCAGAGAGGCGCCACGACCGAGTGGACCTTACAGATCTTACAGAGGCCCTTCAGAGACC	1717		
Db	1495	TCCAGAGAGGCGCCACGAGGCTGAGGCTTACAGATCTTACAGAGGCCCTTCAGAGAAC	1554		
Qy	1718	TGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCAACGACGTGAAGCAGCTGA	1777		
Db	1555	TGAAGACCGGCAAGTACGCCCGCATGCGCGGCCACACCAACGACGTGAAGCAGCTGA	1614		
Qy	1778	CCGAGGCCGTGCAAGAGATCGCCATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGT	1837		
Db	1615	CCGAGGCCGTGCAAGAGTGAACACCGAGACATCGTGATCTGGGGCAAGATCTCCCAAGT	1674		
Qy	1838	TCGCGCTGCCCATCCAGAGAGACCTGGGAGACCTGTGTGACCGATCTTGGCAGGCCA	1897		
Db	1675	TCAAGCTGCCCATCCAGAGAGAGACCTGGGAGGCTGTGTGATCTGGGAGGCTTGGCAGGCCA	1734		
Qy	1898	CCTGGATCCCGAGTGGGAGTTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTGATCCAGC	1957		
Db	1735	CCTGGATCCCGAGTGGGAGTTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTGATCCAGC	1794		
Qy	1958	TGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCTAGTGGAGCGGCGCCCAACCGCG	2017		
Db	1795	TGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCTAGTGGAGCGGCGCCCAACCGCG	1854		
Qy	2018	AGACCAAGATCGGCAAGGCCGCTTACGTGACCGACCGGGGGCGGAGAGATCTGTAGCC	2077		
Db	1855	AGACCAAGCTGGGCGAAGGCCGCTTACGTGACCGACCGGGGGCGGAGAGGTTGTAGCA	1914		
Qy	2078	TGACCGGAGACACCAACAGAGACCGAGCTGACAGGCCATCCAGCTGCGCCCTGCAGAGCA	2137		
Db	1915	TGCGCGACACCAACAGAGACCGAGCTGACAGGCCATCCACCTGGCCCTTCAGAGNCA	1974		
Qy	2138	CGCGCAGCGAGGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCGATCATCCAGGCC	2197		
Db	1975	CGGCGCTGGAGGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCGATCATCCAGGCC	2034		
Qy	2198	AGCCCGACAGAGGAGGAGCGAGCTGTGTGAACAGATCATCGAGCAGCTGATCAAGAGG	2257		
Db	2035	AGCCCGACAGAGGAGGAGGAGCTGTGTGAGCCAGATCATCGAGCAGCTGATCAAGAGG	2094		
Qy	2258	AGAAGGTGATCTGAGCTGGGTGCCCGCCACCAAGGGCATCGGCGGCAACGAGCAGATCG	2317		
Db	2095	AGAAGGTGATCTGAGCTGGGTGCCCGCCACCAAGGGCATCGGCGGCAACGAGCAGGTGG	2154		



Qy	1778	CCGAGGCGCTGCAGAGATCGCCATGGAGAGATCGTGATCTGGGGCAAGACCCCCCAAGT	1837
Db	1615		1674
Qy	1838	TCCGCTGCGCCATCCAGAAAGAGACCTGGGAGACCTGGTGACCGACTACTGGCAGGCGCA	1897
Db	1675		1734
Qy	1898	CCTGGATCCCGAGTGGAGTTCTGTGAACACCCCGCCCTCGTGAAGCTGTGTACCAAGC	1957
Db	1735		1794
Qy	1958	TGGAGAGAGCCCATCATCGCGCGCGAGACCTTCTAGTGTGAACCGCGCCGCCAACCGCG	2017
Db	1795		1854
Qy	2018	AGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGCGCAGAAAGATCGTGAGCC	2077
Db	1855		1914
Qy	2078	TGACCGAGACCAACCAACAGAAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA	2137
Db	1915		1974
Qy	2138	GCGGAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCC	2197
Db	1975		2034
Qy	2198	AGCCCGACAAGCGAGCGAGCTGGTGAAACCAAGATCATCGAGCAGCTGATCAAGAAGG	2257
Db	2035		2094
Qy	2258	AGAAAGGTGTACTTGAGCTGGGTGCCCGCCCAAGGGCATCGGGCGCAACGAGCAGATCG	2317
Db	2095		2154
Qy	2318	ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTTGAGCGGCATCGATGGCGGCA	2377
Db	2155		2214
Qy	2378	TCGTGATCTACCACTACATGACGACCTGTACGTGGCGAGCGCGCCCTAGGATCGATT	2437
Db	2215		2274
Qy	2438	AAAAGCTTCCCGGGCTAGCACCGGTGAATTC	2469
Db	2275		2306

Search completed: June 1, 2005, 20:20:25  
Job time : 10578.3 secs



***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 10:29:57 ; Search time 1414.77 Seconds  
(without alignments)  
10728.225 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469

Sequence: 1 gtcagccaccatggccga.....gggctagaccgtgaattc 2469

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5706582 seqs, 3073711274 residues

Total number of hits satisfying chosen parameters: 11413164

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications NA:\*
- 1: /cgn2\_6/ptodata/1/pubpna/PCT NEW PUB.seq.\*
  - 2: /cgn2\_6/ptodata/1/pubpna/PCT NEW PUB.seq.\*
  - 3: /cgn2\_6/ptodata/1/pubpna/US06 NEW PUB.seq.\*
  - 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
  - 5: /cgn2\_6/ptodata/1/pubpna/US07 NEW PUB.seq.\*
  - 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
  - 7: /cgn2\_6/ptodata/1/pubpna/US08 NEW PUB.seq.\*
  - 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
  - 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
  - 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
  - 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
  - 12: /cgn2\_6/ptodata/1/pubpna/US09 NEW PUB.seq.\*
  - 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
  - 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
  - 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
  - 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
  - 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
  - 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
  - 19: /cgn2\_6/ptodata/1/pubpna/US10 NEW PUB.seq.\*
  - 20: /cgn2\_6/ptodata/1/pubpna/US11 NEW PUB.seq.\*
  - 21: /cgn2\_6/ptodata/1/pubpna/US60 NEW PUB.seq.\*
  - 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2469	100.0	2469	10	US-09-899-575-30
2	2457	99.5	2457	15	US-10-190-435-45
3	2457	99.5	2457	16	US-10-190-305A-39
4	2442.2	98.9	2463	10	US-09-899-575-31
5	2434.6	98.6	2457	15	US-10-190-435-44
6	2434.6	98.6	2457	16	US-10-190-305A-38
7	2415.4	97.8	2457	10	US-09-899-575-32
8	2401.8	97.3	2445	15	US-10-190-435-43
9	2401.8	97.3	2445	16	US-10-190-305A-37
10	2394.8	97.0	3930	15	US-10-190-435-9
11	2393.2	96.9	3930	15	US-10-190-435-10

12	2393.2	96.9	3930	15	US-10-190-435-11	Sequence 11, Appl
13	2393.2	96.9	5184	15	US-10-190-435-58	Sequence 58, Appl
14	2393.2	96.9	5184	16	US-10-190-305A-82	Sequence 82, Appl
15	2362.8	95.6	3531	15	US-10-190-435-13	Sequence 13, Appl
16	2361.2	95.6	3537	15	US-10-190-435-14	Sequence 14, Appl
17	2361.2	95.6	3537	15	US-10-190-435-15	Sequence 15, Appl
18	2360.2	95.6	5145	15	US-10-190-435-12	Sequence 12, Appl
19	2360.2	95.6	5145	16	US-10-190-305A-12	Sequence 12, Appl
20	2349.4	95.2	3624	15	US-10-190-435-47	Sequence 47, Appl
21	2349.4	95.2	3624	16	US-10-190-305A-41	Sequence 41, Appl
22	2301.8	93.2	3607	15	US-10-190-435-48	Sequence 48, Appl
23	2301.8	93.2	3607	16	US-10-190-305A-42	Sequence 42, Appl
24	2283.6	92.5	3597	15	US-10-190-435-46	Sequence 46, Appl
25	2283.6	92.5	3597	16	US-10-190-305A-40	Sequence 40, Appl
26	2149.6	87.1	2472	16	US-10-241-009-32	Sequence 32, Appl
27	2149.6	87.1	2472	16	US-10-190-434B-32	Sequence 32, Appl
28	2149.6	87.1	2472	16	US-10-190-305A-34	Sequence 34, Appl
29	2121.2	85.9	2466	16	US-10-241-009-31	Sequence 31, Appl
30	2121.2	85.9	2466	16	US-10-190-434B-31	Sequence 31, Appl
31	2121.2	85.9	2466	16	US-10-190-305A-33	Sequence 33, Appl
32	2094.4	84.8	2460	16	US-10-241-009-30	Sequence 30, Appl
33	2094.4	84.8	2460	16	US-10-190-434B-30	Sequence 30, Appl
34	2094.4	84.8	2460	16	US-10-190-305A-32	Sequence 32, Appl
35	2093.4	84.8	3564	16	US-10-241-009-13	Sequence 13, Appl
36	2093.4	84.8	3564	16	US-10-241-009-14	Sequence 14, Appl
37	2093.4	84.8	3564	16	US-10-190-434B-13	Sequence 13, Appl
38	2093.4	84.8	3564	16	US-10-190-434B-14	Sequence 14, Appl
39	2092.8	84.8	4716	15	US-10-190-435-17	Sequence 17, Appl
40	2092.8	84.8	4716	16	US-10-190-305A-13	Sequence 13, Appl
41	2089	84.6	3999	16	US-10-241-009-9	Sequence 9, Appl
42	2089	84.6	3999	16	US-10-190-434B-9	Sequence 9, Appl
43	2087.4	84.5	3999	16	US-10-241-009-10	Sequence 10, Appl
44	2087.4	84.5	3999	16	US-10-241-009-11	Sequence 11, Appl
45	2087.4	84.5	3999	16	US-10-190-434B-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-899-575-30  
; Sequence 30, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megede, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Estrelita Janse.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; PRIOR FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 09/475,704  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 2469  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PR975(+)  
US-09-899-575-30

Query Match 100.0%; Score 2469; DB 10; Length 2469;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCAGCCACCACCTGGCCGAGCCATGAGCCAGCCAGCCAGCCACATCTGATGCAG 60

Db 1 GTCAGCCACCACCTGGCCGAGCCATGAGCCAGCCAGCCAGCCACATCTGATGCAG 60

Qy	61	CGAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGC	120
Db	61	CGAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGC	120
Qy	121	CACATGCGCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGGAG	180
Db	121	CACATGCGCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGGAG	180
Qy	181	GGCCACAGATGAAGACTGCAACGAGCGCCAGCGCCAACTTCTTCGCGAGGACCTGGCC	240
Db	181	GGCCACAGATGAAGACTGCAACGAGCGCCAGCGCCAACTTCTTCGCGAGGACCTGGCC	240
Qy	241	TTCCCCAGGCAAGGCCCCCGAGTTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Db	241	TTCCCCAGGCAAGGCCCCCGAGTTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Qy	301	AGCCGCGAGTGCAGGTGCGCGCGCAACACCCCGCAGCGAGCGCCGCGCGAGCGCCAG	360
Db	301	AGCCGCGAGTGCAGGTGCGCGCGCAACACCCCGCAGCGAGCGCCGCGCGAGCGCCAG	360
Qy	361	GGCACCTGAACTTCCCGCAGATCACTCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG	420
Db	361	GGCACCTGAACTTCCCGCAGATCACTCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG	420
Qy	421	GGCGGCCAGATCAAGGAGGCCCTGTGGAACACCGGCGCCGACACACCGTGTGGAGGAG	480
Db	421	GGCGGCCAGATCAAGGAGGCCCTGTGGAACACCGGCGCCGACACACCGTGTGGAGGAG	480
Qy	481	ATGAGCCTGCGCGCAAGTGGAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG	540
Db	481	ATGAGCCTGCGCGCAAGTGGAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG	540
Qy	541	GTGCGCGAGTACACACGATCTGTATGAGATCTGCGGCAAGAGGCGCATCGGCAACGTTG	600
Db	541	GTGCGCGAGTACACACGATCTGTATGAGATCTGCGGCAAGAGGCGCATCGGCAACGTTG	600
Qy	601	CTCATCGGCCCCACCCCGTGAACATCATCGGCGGCAACATGCTGACCCAGCTGGGCTGC	660
Db	601	CTCATCGGCCCCACCCCGTGAACATCATCGGCGGCAACATGCTGACCCAGCTGGGCTGC	660
Qy	661	ACCCTGAACTTCCCATCAGCCCATCGAGACCTGCGCGTGAAGCTGAAGCCCGGCGATG	720
Db	661	ACCCTGAACTTCCCATCAGCCCATCGAGACCTGCGCGTGAAGCTGAAGCCCGGCGATG	720
Qy	721	GACGCCCCAAGTGAAGCTGAGAGTGGCCCTTACCGAGAGAGATCAAGGCGCTGACCGCC	780
Db	721	GACGCCCCAAGTGAAGCTGAGAGTGGCCCTTACCGAGAGAGATCAAGGCGCTGACCGCC	780
Qy	781	ATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCAACCAAGATCGGCCCCGAGAAACCCCTAC	840
Db	781	ATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCAACCAAGATCGGCCCCGAGAAACCCCTAC	840
Qy	841	AACACCCCGTGTTCGCGCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Db	841	AACACCCCGTGTTCGCGCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Qy	901	TTCCGCGAGTGAACCAAGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCGATCCCCCAC	960
Db	901	TTCCGCGAGTGAACCAAGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCGATCCCCCAC	960
Qy	961	CCGCGCGCTGAAGAGAGAGAGCGTCAACCGTGTGACCTGGCGAGCGCTACTTC	1020
Db	961	CCGCGCGCTGAAGAGAGAGAGCGTCAACCGTGTGACCTGGCGAGCGCTACTTC	1020
Qy	1021	AGGCTGCCCCGAGGAGCTTCCGCAAGTACACCGCCCTTACCATTCCCGAGCATCAAC	1080
Db	1021	AGGCTGCCCCGAGGAGCTTCCGCAAGTACACCGCCCTTACCATTCCCGAGCATCAAC	1080
Qy	1081	AACGAGACCCCGGCGATCCCGTACAGTAACAAAGTGTGCCCCAGGGCTGGAGGGGAGC	1140
Db	1081	AACGAGACCCCGGCGATCCCGTACAGTAACAAAGTGTGCCCCAGGGCTGGAGGGGAGC	1140
Qy	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTTCGCGCGCCCAACCCC	1200

Db	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTTCGCGCGCCCAACCCC	1200
Qy	1201	GAGATCGTGTATCTACAGTACATGGAGGACCTGTACGTGGGAGCGACCTGGAGATCGGC	1260
Db	1201	GAGATCGTGTATCTACAGTACATGGAGGACCTGTACGTGGGAGCGACCTGGAGATCGGC	1260
Qy	1261	CAGCACCGCGCAAGATCGAGGAGTGCAGCAAGACCTGCTGCGCTGGGCTTTCACCAAC	1320
Db	1261	CAGCACCGCGCAAGATCGAGGAGTGCAGCAAGACCTGCTGCGCTGGGCTTTCACCAAC	1320
Qy	1321	CCCGACAAGAACCAACAGAGGAGCCCTCTCTGTGTGATGGGCTTACGAGCTGCACCCC	1380
Db	1321	CCCGACAAGAACCAACAGAGGAGCCCTCTCTGTGTGATGGGCTTACGAGCTGCACCCC	1380
Qy	1381	GACAACTGAGCGTGCAGCCCATCGAGCTGCGCGAGAGGAGCTGGACCCGTGAACGAC	1440
Db	1381	GACAACTGAGCGTGCAGCCCATCGAGCTGCGCGAGAGGAGCTGGACCCGTGAACGAC	1440
Qy	1441	ATCCGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG	1500
Db	1441	ATCCGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG	1500
Qy	1501	CGCCAGCTGTGCAAGCTGTGCGCGCCCAAGGCCCTTGACCAATCTGCTGCCCTGACC	1560
Db	1501	CGCCAGCTGTGCAAGCTGTGCGCGCCCAAGGCCCTTGACCAATCTGCTGCCCTGACC	1560
Qy	1561	GAGGAGCGAGCTGAGCTGCGCGAGAACCGGAGATCTCTGCGCGAGCCCGTGCACGGC	1620
Db	1561	GAGGAGCGAGCTGAGCTGCGCGAGAACCGGAGATCTCTGCGCGAGCCCGTGCACGGC	1620
Qy	1621	GTGTACTAGACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAAAGCAGGCGCACGACCAG	1680
Db	1621	GTGTACTAGACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAAAGCAGGCGCACGACCAG	1680
Qy	1681	TGGACCTTACAGATCTTACAGAGGCGCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAG	1740
Db	1681	TGGACCTTACAGATCTTACAGAGGCGCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAG	1740
Qy	1741	ATGCGCACCGCCCAACCAACGAGCTGAAGCAGCTGACCGGCGCGTGCAGAGATCGCC	1800
Db	1741	ATGCGCACCGCCCAACCAACGAGCTGAAGCAGCTGACCGGCGCGTGCAGAGATCGCC	1800
Qy	1801	ATGCGAGCATCTGTATCTGGGCAAGACCCCAAGTTCGCTGCCCTCATTCAGAAAGGAG	1860
Db	1801	ATGCGAGCATCTGTATCTGGGCAAGACCCCAAGTTCGCTGCCCTCATTCAGAAAGGAG	1860
Qy	1861	ACCTGGGAGACTGTGTGGAACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTC	1920
Db	1861	ACCTGGGAGACTGTGTGGAACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTC	1920
Qy	1921	GTGAAACACCCCGCTGTGTGAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGGC	1980
Db	1921	GTGAAACACCCCGCTGTGTGAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGGC	1980
Qy	1981	GCCGAGACTTCTACCTGGAAGCGCGCCCAACCGGAGACCAAGATTCGGCAAGGCCGCGC	2040
Db	1981	GCCGAGACTTCTACCTGGAAGCGCGCCCAACCGGAGACCAAGATTCGGCAAGGCCGCGC	2040
Qy	2041	TACGTGACCGACCCGGGCGCGAGAGATCGTGAGCTTGACCGAGACCAACCAAGAG	2100
Db	2041	TACGTGACCGACCCGGGCGCGAGAGATCGTGAGCTTGACCGAGACCAACCAAGAG	2100
Qy	2101	ACCGAGTGCAGGCGCATTCAGCTGGGCCCTGCGAGGACAGCGGCGAGGAGTGAACATCGTG	2160
Db	2101	ACCGAGTGCAGGCGCATTCAGCTGGGCCCTGCGAGGACAGCGGCGAGGAGTGAACATCGTG	2160
Qy	2161	ACCGAGCGAGTACCGCTTGGGCTATCATCCAGCGCCAGCCCGACAAAGAGCGAGGAGCAG	2220
Db	2161	ACCGAGCGAGTACCGCTTGGGCTATCATCCAGCGCCAGCCCGACAAAGAGCGAGGAGCAG	2220
Qy	2221	CTGGTGAACCCAGATCATTCAGAGCAGCTGATCAAGAGGAGAGGTGTACTTGAAGTGGGTG	2280
Db	2221	CTGGTGAACCCAGATCATTCAGAGCAGCTGATCAAGAGGAGAGGTGTACTTGAAGTGGGTG	2280

Db 2221 CTGGTGAACAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTG 2280  
Qy 2281 CCCGCCCAAGGGCATCGCGGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGCATC 2340  
Db 2281 CCCGCCCAAGGGCATCGCGGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGCATC 2340  
Qy 2341 CGCAAGTGTCTTCTCGGAGCGGATCGATCGCGGCAATCGTGATCTACCAAGTACATGAC 2400  
Db 2341 CGCAAGTGTCTTCTCGGAGCGGATCGATCGCGGCAATCGTGATCTACCAAGTACATGAC 2400  
Qy 2401 GACCTGTACGTGGGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460  
Db 2401 GACCTGTACGTGGGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460  
Qy 2461 GGTGAATTC 2469  
Db 2461 GGTGAATTC 2469

RESULT 2  
US-10-190-435-45  
; Sequence 45, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MESEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBERG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190.435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Poliopt\_C  
US-10-190-435-45

Query Match 99.5%; Score 2457; DB 15; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GCCACCATGGCCGAGGCGCATGAGCAGGCCACGAGCGCCAAACATCTCTGATGCGAGCGAGC 66  
Db 1 GCCACCATGGCCGAGGCGCATGAGCAGGCCACGAGCGCCAAACATCTCTGATGCGAGCGAGC 60  
Qy 67 AACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGGCCACATC 126  
Db 61 AACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGGCCACATC 120  
Qy 127 GCCGCGAACTGCGCGGCCCGCCGCAAGAAGGGCTGCTGGAGTGGCGGCAAGAGGGCCAC 186  
Db 121 GCCGCGAACTGCGCGGCCCGCCGCAAGAAGGGCTGCTGGAGTGGCGGCAAGAGGGCCAC 180  
Qy 187 CAGATGAAGAGTGTGACCGAGCGCGCAGGCCAACTTCTTCGCGAGGAGACTTGGCCTTCC 246  
Db 181 CAGATGAAGAGTGTGACCGAGCGCGCAGGCCAACTTCTTCGCGAGGAGACTTGGCCTTCC 240  
Qy 247 CAGGCGAAGGCGCGGAGTTCCTCCAGCGAGCAGAGAACCGCGCAACAGCCACCGAGCGCG 306  
Db 241 CAGGCGAAGGCGCGGAGTTCCTCCAGCGAGCAGAGAACCGCGCAACAGCCACCGAGCGCG 300  
Qy 307 GAGCTGAGGTGCGCGGCGCAACCCCGCAGCGAGGCGCGGCGCGGAGCGCGAGCGGCGAC 366  
Db 301 GAGCTGAGGTGCGCGGCGCAACCCCGCAGCGAGGCGCGGCGCGGAGCGCGAGCGGCGAC 360  
Qy 367 CTGAACCTTCCCGCAGATCACTCTGTGGAGCGCGCCCTGTGTGAGCATCAAGGTGGCGGC 426

Db 361 CTGAACCTTCCCGCAGATCACTCTGTGGAGCGCGCCCTGTGTGAGCATCAAGGTGGCGGC 420  
Qy 427 CAGATCAAGGAGGCGCTTGTGGACACCGGCGCGCGACGACACCGTGTCTGGAGGAGATGAGC 486  
Db 421 CAGATCAAGGAGGCGCTTGTGGACACCGGCGCGCGACGACACCGTGTCTGGAGGAGATGAGC 480  
Qy 487 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGCGCGCTTTCATCAAGGTGGCG 546  
Db 481 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGCGCGCTTTCATCAAGGTGGCG 540  
Qy 547 CAGTACGACCAAGATCTCTGATCGAGATCTGGCGCAAGAGGCGCATCGCGCATCGTCTGATC 606  
Db 541 CAGTACGACCAAGATCTCTGATCGAGATCTGGCGCAAGAGGCGCATCGCGCATCGTCTGATC 600  
Qy 607 GGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666  
Db 601 GGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660  
Qy 667 AACTTCCCATCAGCCCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGCGATGAGCGGC 726  
Db 661 AACTTCCCATCAGCCCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGCGATGAGCGGC 720  
Qy 727 CCCAAGGTGAAGCAGTGGCGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
Db 721 CCCAAGGTGAAGCAGTGGCGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 780  
Qy 787 GAGGAGATGGAGAAGGAGGCGCAAGATCAACCAAGATCGGCCCCCGAGAAACCCCTCAACAACC 846  
Db 781 GAGGAGATGGAGAAGGAGGCGCAAGATCAACCAAGATCGGCCCCCGAGAAACCCCTCAACAACC 840  
Qy 847 CCCGTGTTCGCCATCAAGAAGAGGAGCAGCACCAGTGGCGCAAGCTGGTGGATCTTCGC 906  
Db 841 CCCGTGTTCGCCATCAAGAAGAGGAGCAGCACCAGTGGCGCAAGCTGGTGGATCTTCGC 900  
Qy 907 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCGAGCTGGGCGATCCGCCACCCCGGC 966  
Db 901 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCGAGCTGGGCGATCCGCCACCCCGGC 960  
Qy 967 GGCCTGAAGAAGAGAGAGCGGTGACCGTGTGGACGTGGGCGCAGCGCTACTTTCAGGGTG 1026  
Db 961 GGCCTGAAGAAGAGAGAGAGCGGTGACCGTGTGGACGTGGGCGCAGCGCTACTTTCAGGGTG 1020  
Qy 1027 CCCCTGAGCAGGACTTTCGCCAAGTACACCGCTTTCACATCTCCCGAGATCAACAAGAG 1086  
Db 1021 CCCCTGAGCAGGACTTTCGCCAAGTACACCGCTTTCACATCTCCCGAGATCAACAAGAG 1080  
Qy 1087 ACCCGCGGATCCGCTACCAAGTACCAAGCTGTCGCCCGAGGGCTGGAAAGGGCAGCCCCAGC 1146  
Db 1081 ACCCGCGGATCCGCTACCAAGTACCAAGCTGTCGCCCGAGGGCTGGAAAGGGCAGCCCCAGC 1140  
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTCTGAGAGCCCTTTCGCGCGCCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTCTGAGAGCCCTTTCGCGCGCCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACAGTACATGAGCAGCTGTGACGCGAGCAGCCTTGGAGATCGGCGAGGAC 1266  
Db 1201 GTGATCTACAGTACATGAGCAGCTGTGAGCAGCAGCTTGGAGAGCAGCCTGGAGATCGGCGAGC 1260  
Qy 1267 CCGCGCAAGATCGAGGAGCTGCGCAGCAGCTGTCGCTGGGGCTTTCACACCCCGGAC 1326  
Db 1261 CCGCGCAAGATCGAGGAGCTGCGCAGCAGCTGTCGCTGGGGCTTTCACACCCCGGAC 1320  
Qy 1327 AAGAAGCAGCAGAGAGGCGCCCTTCTCTGTGATGGGTACGAGCTGCAACCCCGCAAG 1386  
Db 1321 AAGAAGCAGCAGAGAGGCGCCCTTCTCTGTGATGGGTACGAGCTGCAACCCCGCAAG 1380  
Qy 1387 TGGAACGCTGAGCCCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAJACGATCCAG 1446  
Db 1381 TGGAACGCTGAGCCCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAJACGATCCAG 1440  
Qy 1447 AAGCTGTGGGCAAGCTGAATCTGGGCGAGCAGATCTACCCCGGCGCATCAAGGTGGCGGC 1506

```

1441 AAGTCGTGGCGAAGCTGAACCTGGCCGAGCCAGATCTACCCCGCATCAAGGTGCGCCAG 1500
1507 CTGTGAAGCTGTGCGCGCGCCAGCCCTGACGACATCTGCCCCCTGACCGAGGAG 1566
1501 CTGTGAAGCTGTGCGCGCGCCAGCCCTGACGACATCTGCCCCCTGACCGAGGAG 1560
1567 GCCAGCTGGAGCTGGCGCGAGAACCCGCGAGATCTGCGCGAGCCCGTGACGCGGTGTAC 1626
1561 GCCAGCTGGAGCTGGCGCGAGAACCCGCGAGATCTGCGCGAGCCCGTGACGCGGTGTAC 1620
1627 TAGACCCCAAGCAAGCACTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACGATGGACC 1686
1621 TAGACCCCAAGCAAGCACTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACGATGGACC 1680
1687 TACCAGATCTACGAGGAGCCCTTCAGAACCTGAGACCGGCAAGTACGCGCAAGATGCGC 1746
1681 TACCAGATCTACGAGGAGCCCTTCAGAACCTGAGAACCCGCGAAGTACGCGCAAGATGCGC 1740
1747 ACCGCCACACCAACGACGCTGAAGCAGCTGACCGAGGCCGTGCAAGAGATCGCCATGGAG 1806
1741 ACCGCCACACCAACGACGCTGAAGCAGCTGACCGAGGCCGTGCAAGAGATCGCCATGGAG 1800
1807 AGCATCGTGTCTGTGGGCAAGACCCCAAGTTCCGCTGCCCCATCCAGAAAGGAGACTGG 1866
1801 AGCATCGTGTCTGTGGGCAAGACCCCAAGTTCCGCTGCCCCATCCAGAAAGGAGACTGG 1860
1867 GAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAAC 1926
1861 GAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAAC 1920
1927 ACCCCCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCTGCGCGCCGAG 1986
1921 ACCCCCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCTGCGCGCCGAG 1980
1987 ACCTTTACGTGGAAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2046
1981 ACCTTTACGTGGAAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040
2047 ACCGACCGGGCGCGGAGAGATCGTGAGCTGACCGAGACCCACCAACGAGAGACCGAG 2106
2041 ACCGACCGGGCGCGGAGAGATCGTGAGCTGACCGAGACCCACCAACGAGAGACCGAG 2100
2107 CTGACGACCATCAAGCTGGCCCTGACGACGAGCGGACGAGGTGAACATCGTGACCGAC 2166
2101 CTGACGACCATCAAGCTGGCCCTGACGACGAGCGGACGAGGTGAACATCGTGACCGAC 2160
2167 AGCCAGTACCCCTGGGCATCATCCAGGCCAGCCCGACGAAGAGCGAGCGAGCTGGTG 2226
2161 AGCCAGTACCCCTGGGCATCATCCAGGCCAGCCCGACGAAGAGCGAGCGAGCTGGTG 2220
2227 AACGAGATCATCGAGAGCTGATCAAGAGGAGAGGTGTACTGTGAGCTGGGTGCCGCC 2286
2221 AACGAGATCATCGAGAGCTGATCAAGAGGAGAGGTGTACTGTGAGCTGGGTGCCGCC 2280
2287 CACAAGGCGCATCGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGCGCATCCGCAAG 2346
2281 CACAAGGCGCATCGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGCGCATCCGCAAG 2340
2347 GTGCTGTCTCGGACCGCATCGATGGCGGCATCGTGATCTACAGTATCATGGAAGACCTG 2406
2341 GTGCTGTCTCGGACCGCATCGATGGCGGCATCGTGATCTACAGTATCATGGAAGACCTG 2400
2407 TAGTGGGACCGCGCGCCCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2463
2401 TAGTGGGACCGCGCGCCCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

```

RESULT 3

```

US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan

```

```

; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

```

```

Query Match      99.5%; Score 2457; DB 16; Length 2457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCACCATGCGCGAGGCCATGAGCCAGGCCACCGCCACCATCTCTGATGCGAGCGCAGC 66
DB 1 GCCACCATGCGCGAGGCCATGAGCCAGGCCACCGCCACCATCTCTGATGCGAGCGCAGC 60
QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATC 126
DB 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATC 120
QY 127 GCCCGCAACTGCGCGCCCCCCCCCAAGAGGGGTGTGTGAAGTGTGCGCAAGGAGGGCCAC 186
DB 121 GCCCGCAACTGCGCGCCCCCCCCCAAGAGGGGTGTGTGAAGTGTGCGCAAGGAGGGCCAC 180
QY 187 CAGATCAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGGCTTCCCC 246
DB 181 CAGATCAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGGCTTCCCC 240
QY 247 CAGGGCAAGGCCCGCGAGTTTCCCCAGCGAGCAGAAACCGGCCCAACAGCCCCACAGCCGC 306
DB 241 CAGGGCAAGGCCCGCGAGTTTCCCCAGCGAGCAGAAACCGGCCCAACAGCCCCACAGCCGC 300
QY 307 GAGCTCAGGTGCGCGCGCGCAAAACCCCGCAGCGCGCGGCCGCCCGCGAGCGCCAGGGGACC 366
DB 301 GAGCTCAGGTGCGCGCGCGCAAAACCCCGCAGCGAGAACCCCGCAGCGCGCGCGAGCGCCAGGGGACC 360
QY 367 CTGAACTTCCCCCAGATCAACCTGTGCGAGCGGCCCTTGTGTGAGCATCAAGGTGGCGGC 426
DB 361 CTGAACTTCCCCCAGATCAACCTGTGTGAGCGGCCCTTGTGTGAGCATCAAGGTGGCGGC 420
QY 427 CAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGACACACCTGTGTGAGGAGATGAGC 486
DB 421 CAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGACACACCTGTGTGAGGAGATGAGC 480
QY 487 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAAGGTGCGC 546
DB 481 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAAGGTGCGC 540
QY 547 CAGTACGACCGAGATCTGTATCGAGATCTGCGGCAAGAGGCCATCGGACCGGTGCTGATC 606
DB 541 CAGTACGACCGAGATCTGTATCGAGATCTGCGGCAAGAGGCCATCGGACCGGTGCTGATC 600
QY 607 GGCCCCACCCCGGTGAACATCATCTGCGCGCAACATGCTGACCCAGGTGGGCTGCACCCCTG 666
DB 601 GGCCCCACCCCGGTGAACATCATCTGCGCGCAACATGCTGACCCAGGTGGGCTGCACCCCTG 660
QY 667 AACTTCCCCATCAGCCCCCATTCGAGACCGGTGCCGTGAAGTGAAGCCCGGCGATGAGCGGC 726
DB 661 AACTTCCCCATCAGCCCCCATTCGAGACCGGTGCCGTGAAGTGAAGCCCGGCGATGAGCGGC 720
QY 727 CCCAAGGTGAAGCAGTGGGCCCTTGACCGGAGGAGAAATCAAGGCCCTGACCGCCCATCTGC 786
DB 721 CCCAAGGTGAAGCAGTGGGCCCTTGACCGGAGGAGAAATCAAGGCCCTGACCGCCCATCTGC 780

```





Qy	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGC	120
Db	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGC	120
Qy	121	CACATCGCCGCAACTCGCGCCCCCGGCGAAGAGGCTGCTGGAAGTGGCGCAAGGAG	180
Db	121	CACATCGCCGCAACTCGCGCCCCCGGCGAAGAGGCTGCTGGAAGTGGCGCAAGGAG	180
Qy	181	GGCCACAGATGAAGGACTGCAACGAGCGCCCAACTTCTTCCGCGAGGACCTGGCC	240
Db	181	GGCCACAGATGAAGGACTGCAACGAGCGCCCAACTTCTTCCGCGAGGACCTGGCC	240
Qy	241	TTTCCCGCAGGCAAGGCCCGGAGTTTCCGAGGAGCAGAAACCGCGCCACAGCCCAAC	300
Db	241	TTTCCCGCAGGCAAGGCCCGGAGTTTCCCGAGGAGCAGAAACCGCGCCACAGCCCAAC	300
Qy	301	AGCCGCGAGCTGCAAGTGCAGGCGGCAACACCCCGCAGCGAGGCGCGCGCAGCGCCAG	360
Db	301	AGCCGCGAGCTGCAAGTGCAGGCGGCAACACCCCGCAGCGAGGCGCGCGCAGCGCCAG	360
Qy	361	GGCACCCTGAACTTCCGCCAGATCACCCCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG	420
Db	361	GGCACCCTGAACTTCCGCCAGATCACCCCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG	420
Qy	421	GGCGGCAGATCAAGGAGGCCCTCTCTGACACCGCGCGCAGACACCGTCTGAGGAG	480
Db	421	GGCGGCAGATCAAGGAGGCCCTCTCTGACACCGCGCGCAGACACCGTCTGAGGAG	480
Qy	481	ATGAGCCTTCCCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGCGGCTTCATCAAG	540
Db	481	ATGAGCCTTCCCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGCGGCTTCATCAAG	540
Qy	541	GTGCGCAGTACGACAGTCTCTGATCGAGATCTGCGGCGAAGAGGCCATCGGCAACG	600
Db	541	GTGCGCAGTACGACAGTCTCTGATCGAGATCTGCGGCGAAGAGGCCATCGGCAACG	600
Qy	601	CTGATCGGCCCCACCCCGTGAAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Db	601	CTGATCGGCCCCACCCCGTGAAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Qy	661	ACCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGTCCCGTGAAGCTGAAGCCCGCATG	720
Db	661	ACCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGTCCCGTGAAGCTGAAGCCCGCATG	720
Qy	721	GACGCCCCAAGGTGAAGCTGTGACCGAGAGAGATCAAGGCCCTGACCGCC	780
Db	721	GACGCCCCAAGGTGAAGCTGTGACCGAGAGAGATCAAGGCCCTGACCGCC	780
Qy	781	ATCTGCGAGGATGGAAGAGGGCAAGATCAACCAAGATCGGCCCCGAGAACCCCTAC	840
Db	781	ATCTGCGAGGATGGAAGAGGGCAAGATCAACCAAGATCGGCCCCGAGAACCCCTAC	840
Qy	841	AACACCCCGTGTTCGCATCAAGAGAGCAGCACCAGTGGCGCGAGCTGGTGAC	900
Db	841	AACACCCCGTGTTCGCATCAAGAGAGCAGCACCAGTGGCGCGAGCTGGTGAC	900
Qy	901	TTTCGCGAGCTGAACAAGCGCACCCAGACTTCTGCGAGGTGCACTGGGCATCCCCCAC	960
Db	901	TTTCGCGAGCTGAACAAGCGCACCCAGACTTCTGCGAGGTGCACTGGGCATCCCCCAC	960
Qy	961	CCCGCGGCTTGAAGAGAGAGAGCGTGCACCGTGTGGAAGTGGCGCAGCTACTTC	1020
Db	961	CCCGCGGCTTGAAGAGAGAGAGCGTGCACCGTGTGGAAGTGGCGCAGCTACTTC	1020
Qy	1021	AGCGTGCCCTGGACGAGGACTTCGCGAAGTACACCGCTTCCACCATCCCGAGCATCAAC	1080
Db	1021	AGCGTGCCCTGGACGAGGACTTCGCGAAGTACACCGCTTCCACCATCCCGAGCATCAAC	1080
Qy	1081	AACGAGACCCCGGCATCCGCTACCAAGTACAAACCGTGTGCCCCCAGGGCTGGAAGGGCAGC	1140
Db	1081	AACGAGACCCCGGCATCCGCTACCAAGTACAAACCGTGTGCCCCCAGGGCTGGAAGGGCAGC	1140

Qy	1141	CCAGCATCTTCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCGCGCCCGCAACCC	1201
Db	1141	CCAGCATCTTCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCGCGCCCGCAACCC	1200
Qy	1201	GAGATCGTGATCTACCAGTACATGAGCGACCTCTGATGGGCGACGACCTGGAGATCGGC	1260
Db	1201	GAGATCGTGATCTACCA-----GGCCCCCTGTATGCTGGGCGAGGACCTGGAGATCGGC	1254
Qy	1261	CAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGGCTGGGCTTCACCAACC	1320
Db	1255	CAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGGCTGGGCTTCACCAACC	1314
Qy	1321	CCCGACAAGAGCACAAGAGAGCCGCCCTTCTGTGTGATGGGCTACGAGCTGCACCC	1380
Db	1315	CCCGACAAGAGCACCAGAAGAGAGCCGCCCTTCTGTGTGATGGGCTACGAGCTGCACCC	1374
Qy	1381	GACAAGTGGACCGTGCAGCCCATCGAGCTCCCGAGAAGGAGAGCTGGACCGGTGAACGAC	1440
Db	1375	GACAAGTGGACCGTGCAGCCCATCGAGCTCCCGAGAAGGAGAGCTGGACCGGTGAACGAC	1434
Qy	1441	ATCCAGAAGCTGTGTGGCAAGCTGAACTGGGGCCAGCCAGATCTTACCCCGCATCAAGGTG	1500
Db	1435	ATCCAGAAGCTGTGTGGCAAGCTGAACTGGGGCCAGCCAGATCTTACCCCGCATCAAGGTG	1494
Qy	1501	CGCCAGCTGTGCAAGCTGTGCGCGGGCCAAAGCCCTGACCGACATCTGTGTCGCCCTTGACC	1560
Db	1495	CGCCAGCTGTGCAAGCTGTGCGCGGGCCAAAGCCCTGACCGACATCTGTGTCGCCCTTGACC	1554
Qy	1561	GAGGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGGAGCCCGGTGCACGGC	1620
Db	1555	GAGGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGCGGAGCCCGGTGCACGGC	1614
Qy	1621	GTGTACTTACGACCCACGAAGGACCTGTGTGGCCGAGATCCAGAAGCAGGGGCCACGACCAG	1680
Db	1615	GTGTACTTACGACCCACGAAGGACCTGTGTGGCCGAGATCCAGAAGCAGGGGCCACGACCAG	1674
Qy	1681	TGGACCTTACAGATCTTACGAGGACGCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG	1740
Db	1675	TGGACCTTACAGATCTTACGAGGACGCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG	1734
Qy	1741	ATGGGACCCGCCACACCAAGCAGGTGAACGAGCTGACGAGGCGCTGCGAGAGATCGCC	1800
Db	1735	ATGGGACCCGCCACACCAAGCAGGTGAACGAGCTGACGAGGCGCTGCGAGAGATCGCC	1794
Qy	1801	ATGGAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGGAG	1860
Db	1795	ATGGAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGGAG	1854
Qy	1861	ACTTGGGAGACTTGTGTGGAACGACTACTGGCAGGCCACTTGGATTCGCCGAGTGGGAGTTC	1920
Db	1855	ACTTGGGAGACTTGTGTGGAACGACTACTGGCAGGCCACTTGGATTCGCCGAGTGGGAGTTC	1914
Qy	1921	GTGNAACCCCCCTGTGTGAGCTGTGTGTACAGCTGGAGAGGAGGCCCATCATCTCGGC	1980
Db	1915	GTGNAACCCCCCTGTGTGAGCTGTGTGTACAGCTGGAGAGGAGGCCCATCATCTCGGC	1974
Qy	1981	GCCGAGACCTTCTACGTGGACCGCGCGCCCAACCGCGAGACCAAGATTCGCAAGCGCCGC	2040
Db	1975	GCCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATTCGCAAGCGCCGC	2034
Qy	2041	TAOBTGACCGACCGGGCCGCGCAAGATCGTGAGCCTTGACCGAGACCAACCAACCAAGAG	2100
Db	2035	TACGTGACCGACCGGGCCGCGCAAGATCGTGAGCCTTGACCGAGACCAACCAACCAAGAG	2094
Qy	2101	ACGAGCTGCAGGCCATTCAGCTTGGCCCTCGAGGACAGCGGACCGAGGTGAACATCTGTG	2160
Db	2095	ACGAGCTGCAGGCCATTCAGCTTGGCCCTCGAGGACAGCGGACCGAGGTGAACATCTGTG	2154
Qy	2161	ACGACAGCCAGATACGCCCTTGGGCATCATCATCGGCCCGACCGCAAGAGCGGAGCGAG	2220
Db	2155	ACGACAGCCAGATACGCCCTTGGGCATCATCATCGGCCCGACCGCAAGAGCGGAGCGAG	2214
Qy	2221	CTGTTGAACAGATCATCGACGAGCTGTATCAAGAAAGGAGAGGTGTACTTGAAGCTGGGTG	2280

```
Db 2215 |||||CTGTGAACAGATCATCGACAGCTGATCAAGAGGAGAGGTTGATCCTGAGCTGGGTG 2274
Qy 2281 CCGCCCAACAGGGCATCGCGGCAACGAGCAGATCGAAGAGCTGGTGAGCAAGGGCATC 2340
Db 2275 CCGCCCAACAGGGCATCGCGGCAACGAGCAGATCGAAGAGCTGGTGAGCAAGGGCATC 2334
Qy 2341 CGAAGGTGCTGTTCTTGGACGGCATCGATCGCGGGCATCTGTGATCTACCAGTACATGGAC 2400
Db 2335 CGAAGGTGCTGTTCTTGGACGGCATCGATCGCGGGCATCTGTGATCTACCAGTACATGGAC 2394
Qy 2401 GACCTGTACGTGGCAGCGGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
Db 2395 GACCTGTACGTGGCAGCGGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2454
Qy 2461 GGTGAATTC 2469
Db 2455 GGTGAATTC 2463
```

## RESULT 5

```
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44
```

```
Query Match 98.6%; Score 2434.6; DB 15; Length 2457;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 1 GTGACGCCACCATGGCGGAGGCCATGAGCCAGGCCACCAGGCCCAACATCCTGATGCAG 60
Db 1 GTGACGCCACCATGGCGGAGGCCATGAGCCAGGCCACCAGGCCCAACATCCTGATGCAG 60

Qy 61 CGCAGCAACTTCAAGGGCCCCAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGGGC 120
Db 61 CGCAGCAACTTCAAGGGCCCCAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGGGC 120

Qy 121 CACATCCCGCGCAACTCCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180
Db 121 CACATCCCGCGCAACTCCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180

Qy 181 GGCCACAGATGAAGAGTGCACCGAGCGGCCCAACTTCTTCCCGAGGACCTTGCC 240
Db 181 GGCCACAGATGAAGAGTGCACCGAGCGGCCCAACTTCTTCCCGAGGACCTTGCC 240

Qy 241 TTCCCCCAGGCAAGGCCCGGAGTTCCCCAGCGAGCAGAACCCGCCCAACAGCCCCACC 300
Db 241 TTCCCCCAGGCAAGGCCCGGAGTTCCCCCAGCGAGCAGAACCCGCCCAACAGCCCCACC 300

Qy 301 AGCCGAGCTGCAGGTGCGCGGACAAACCCCGCAGCGAGGCCGCGCGCGCGCAG 360
Db 301 AGCCGAGCTGCAGGTGCGCGGACAAACCCCGCAGCGAGGCCGCGCGCGCGCAG 360
```

```
Qy 361 GGCAACCTGAACCTTCCCCCAGATCACTCTGGCAGAGCGCCCCCTGGTGAGCATCAAGGTG 420
Db 361 GGCAACCTGAACCTTCCCCCAGATCACTCTGGCAGAGCGCCCCCTGGTGAGCATCAAGGTG 420

Qy 421 GGCGGCCAGATCAAGGAGGCCCTCTGTGACACCGCGCGCCGACGACACCGTGTCTGAGGAG 480
Db 421 GGCGGCCAGATCAAGGAGGCCCTCTGTGACACCGCGCGCCGACGACACCGTGTCTGAGGAG 480

Qy 481 ATGAGCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540
Db 481 ATGAGCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540

Qy 541 GTGCGCAGTACACGACAGATCTCTGATCGAGATCTCGCGCAGAAGGCCATCGGCACCGTG 600
Db 541 GTGCGCAGTACACGACAGATCTCTGATCGAGATCTCGCGCAGAAGGCCATCGGCACCGTG 600

Qy 601 CTGATCGCGCCCAACCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGCGCCCAACCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy 661 ACCCTGAATCTTCCCATCAGCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGCATG 720
Db 661 ACCCTGAATCTTCCCATCAGCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGCATG 720

Qy 721 GACGCCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780
Db 721 GACGCCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC 780

Qy 781 ATCTGCGAGGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTTAC 840
Db 781 ATCTGCGAGGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTTAC 840

Qy 841 AACACCCCCGTGTTTCGCCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCCGTGTTTCGCCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900

Qy 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGCGGAGGTGCAAGTGGGATCCGCCAC 960
Db 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGCGGAGGTGCAAGTGGGATCCGCCAC 960

Qy 961 CCGCGCGCTGGAAGAAAGAGAGCGTGACCGTGTGACGCGTGGCGAGCGCTACTTTC 1020
Db 961 CCGCGCGCTGGAAGAAAGAGAGCGTGACCGTGTGACGCGTGGCGAGCGCTACTTTC 1020

Qy 1021 AGCGTGCCCTTGGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCTTGGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080

Qy 1081 AACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCTGCGCGGAGGAGGAGGAGG 1140
Db 1081 AACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCTGCGCGGAGGAGGAGGAGG 1140

Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCCCGCGCCGCAACCCC 1200
Db 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCCCGCGCCGCAACCCC 1200

Qy 1201 GAGATCGTGATCTAACAAGTACATGGACGACTGTGCTGGGCGAGCGACTTGGAGATCGGC 1260
Db 1201 GAGATCGTGATCTAACAAGTACATGGACGACTGTGCTGGGCGAGCGACTTGGAGATCGGC 1260

Qy 1261 CAGCAGCCGCAAGATCGAGGAGCTGCGCAAGCAGCTGTGCGCTGGGCTTCAACACC 1320
Db 1261 CAGCAGCCGCAAGATCGAGGAGCTGCGCAAGCAGCTGTGCGCTGGGCTTCAACACC 1320

Qy 1321 CCGCAGAAGAGCAGCAGAGGAGGCCCTTCTGTTGATGGCTACGAGCTACGAGCTGCACCCC 1380
Db 1321 CCGCAGAAGAGCAGCAGAGGAGGCCCTTCTGTTGATGGCTACGAGCTACGAGCTGCACCCC 1380

Qy 1381 GACAAGTGAACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGAACCGTGAACGAC 1440
Db 1381 GACAAGTGAACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGAACCGTGAACGAC 1440

Qy 1441 ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGCATCAAGGTG 1500
Db 1441 ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGCATCAAGGTG 1500
```



QY 721 GACGGCCCAAGGTGAAGCAGTGGCCCTCTGACCGAGGAGAAAGATCAAGGCCCTTCAACCGCC 780  
Db 721 GACGGCCCAAGGTGAAGCAGTGGCCCTCTGACCGAGGAGAAAGATCAAGGCCCTTCAACCGCC 780  
QY 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCGCCCGAAGCCCTTAC 840  
Db 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCGCCCGAAGCCCTTAC 840  
QY 841 AACACCCCGCTGTTCCGCATCAAGAAGAGACAGCACCAAGTGGCGCAAGCTGTTGGAC 900  
Db 841 AACACCCCGCTGTTCCGCATCAAGAAGAGACAGCACCAAGTGGCGCAAGCTGTTGGAC 900  
QY 901 TTCCGCGAGCTGAACAAGCGACCCAGAGATTCTTGGAGAGTGCAGCTGGGCAATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAGCGACCCAGAGATTCTTGGAGAGTGCAGCTGGGCAATCCCCAC 960  
QY 961 CCCGCGGCTGGAAGAGAGAGCGTGAACGCTGTCGAGCGTGGCGAGCGCTACTTC 1020  
Db 961 CCCGCGGCTGGAAGAGAGAGCGTGAACGCTGTCGAGCGTGGCGAGCGCTACTTC 1020  
QY 1021 AGCGTGGCCCTGGACGAGGATTCGCGAAGTACACCGGCTTTCAGCATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGGCCCTGGACGAGGATTCGCGAAGTACACCGGCTTTCAGCATCCCCAGCATCAAC 1080  
QY 1081 AACGAGACCCCGGATCCGCTAACAAGTACAACTGTCGCCCGAGGCTGGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGGATCCGCTAACAAGTACAACTGTCGCCCGAGGCTGGAAGGGCAGC 1140  
QY 1141 CCCAGCATCTCCAGAGCAGATCAACCAAGATCTGGAGCCCTTCCGCGCCGCAACCCC 1200  
Db 1141 CCCAGCATCTCCAGAGCAGATCAACCAAGATCTGGAGCCCTTCCGCGCCGCAACCCC 1200  
QY 1201 GAGATCTGTATACCAAGTACAGAGCAGCTGTGAGCAGCAGCTGGAGTGGC 1260  
Db 1201 GAGATCTGTATACCA-----GGCCCCCTGTAGTGGGACGAGCTGGAGTGGC 1260  
QY 1261 CAGCACCGGCAAGATCGAGAGCTGCGAAGCACTGTGCGCTGGGGCTTCAACACC 1320  
Db 1261 CAGCACCGGCAAGATCGAGAGCTGCGAAGCACTGTGCGCTGGGGCTTCAACACC 1320  
QY 1321 CCCGACAGAGCAGCAGAGGAGCCCTTCTGTGTGATGGCTAGGAGCTGACCCC 1380  
Db 1321 CCCGACAGAGCAGCAGAGGAGCCCTTCTGTGTGATGGCTAGGAGCTGACCCC 1380  
QY 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGCACCGTGAACGAC 1440  
Db 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGCACCGTGAACGAC 1440  
QY 1441 ATCCAGAAGCTGTGGGCAAGCTGAATGGGSCCAGCAGATCTACCCCGGCATCAAGGTG 1500  
Db 1441 ATCCAGAAGCTGTGGGCAAGCTGAATGGGSCCAGCAGATCTACCCCGGCATCAAGGTG 1500  
QY 1501 CGCAGCTGTGCAAGCTGTGCGCGGCGCCAGGCGCTTACCGACATCTGTCCTGACC 1560  
Db 1501 CGCAGCTGTGCAAGCTGTGCGCGGCGCCAGGCGCTTACCGACATCTGTCCTGACC 1560  
QY 1561 GAGGAGCCGAGCTGGAGCTGGCGAGAACCGCAGATCTGCGCGAGCCCGTGCACGGC 1620  
Db 1561 GAGGAGCCGAGCTGGAGCTGGCGAGAACCGCAGATCTGCGCGAGCCCGTGCACGGC 1620  
QY 1621 GTGTACTACGACCCAGCAAGGACTGTGGTGGCGGAGATCCAGAAAGAGGCGCCAGCAG 1680  
Db 1621 GTGTACTACGACCCAGCAAGGACTGTGGTGGCGGAGATCCAGAAAGAGGCGCCAGCAG 1680  
QY 1681 TGGACCTACCATCTTACGAGGAGCCCTTCAAGAACTGAGACCGGCAAGTACGCCAAG 1740  
Db 1681 TGGACCTACCATCTTACGAGGAGCCCTTCAAGAACTGAGACCGGCAAGTACGCCAAG 1740  
QY 1741 ATCGGACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCC 1800  
Db 1741 ATCGGACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCC 1800  
QY 1775 ATGGCACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCC 1794  
Db 1775 ATGGCACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCC 1794

QY 1801 ATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTTCCGCTGCCCATCCAGAAGGAG 1860  
Db 1795 ATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTTCCGCTGCCCATCCAGAAGGAG 1854  
QY 1861 ACCTGGGAGACCTGTGGACCGACTACTTGGCAGGCGCACTGTGATCCCGAGTGGGAGTTTC 1920  
Db 1855 ACCTGGGAGACCTGTGTGACCGACTACTTGGCAGGCGCACTGTGATCCCGAGTGGGAGTTTC 1914  
QY 1921 GTGAACACCCCCCTGTGGTGAAGCTGTGTGTACAGCTTGGAGAGGAGCCCATCATCGGC 1980  
Db 1915 GTGAACACCCCCCTGTGGTGAAGCTGTGTGTACAGCTTGGAGAGGAGCCCATCATCGGC 1974  
QY 1981 GCCGAGACCTTCTACGTGAGCGCGCGCCCAACCGCGAGACAAGATCGGCAAGGCCGCGC 2040  
Db 1975 GCCGAGACCTTCTACGTGAGCGCGCGCCCAACCGCGAGACAAGATCGGCAAGGCCGCGC 2034  
QY 2041 TACGTGACCGACCGGGGCGGCGAGAGATCTGTGAGCCTTGACCGAGACCAACCAAG 2100  
Db 2035 TACGTGACCGACCGGGGCGGCGAGAGATCTGTGAGCCTTGACCGAGACCAACCAAG 2094  
QY 2101 ACCGAGCTGCAGGCGCATCCAGCTGGCCCTGTCAGGACACGCGCAGCGAGTGAACATCGTG 2160  
Db 2095 ACCGAGCTGCAGGCGCATCCAGCTGGCCCTGTCAGGACACGCGCAGCGAGTGAACATCGTG 2154  
QY 2161 ACCGAGACCGATGAGCGCTGGGCGATCATCAGGCGCCAGCGCCGAGAGCGAGCGAG 2220  
Db 2155 ACCGAGACCGATGAGCGCTGGGCGATCATCAGGCGCCAGCGCCGAGAGCGAGCGAG 2214  
QY 2221 CTGTGTGAACCAAGTATCATCGAGCAGCTGATCAAGAGAGAGAGAGTGTACCTGAGCTGGGTG 2280  
Db 2215 CTGTGTGAACCAAGTATCATCGAGCAGCTGATCAAGAGAGAGAGAGTGTACCTGAGCTGGGTG 2274  
QY 2281 CCCGCCACCAAGGCGCATCGCGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGCATC 2340  
Db 2275 CCCGCCACCAAGGCGCATCGCGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGCATC 2334  
QY 2341 CGCAAGTGTCTTCTTGGAGCGGATCGATGGCGGCGATCTGATCTACAGTACATGAGC 2400  
Db 2335 CGCAAGTGTCTTCTTGGAGCGGATCGATGGCGGCGATCTGATCTACAGTACATGAGC 2394  
QY 2401 GACCTGTACGTGGGCGCGCGCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACC 2460  
Db 2395 GACCTGTACGTGGGCGCGCGCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACC 2454  
QY 2461 GGT 2463  
Db 2455 GGT 2457

## RESULT 7

US-09-899-575-32  
; Sequence 32, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megede, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Batrelita Janse  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899, 575  
; CURRENT FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 09/475, 704  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 32  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PR975YMMW

US-09-899-575-32

Query Match		97.8%;	Score 2415.4;	DB 10;	Length 2457;		
Best Local Similarity		99.3%;	Pred. No. 0;				
Matches 2451;		Conservative	0;	Mismatches	6;	Indels	12; Gaps 2;
Qy	1	GTGAGGCCACCATGGCCGAGGCGCATGAGCAGGCCACCAAGCCCAACATCTCTGATGCGAG	60				
Db	1	GTGAGGCCACCATGGCCGAGGCGCATGAGCAGGCCACCAAGCCCAACATCTCTGATGCGAG	60				
Qy	61	CGCAGCAATTCGAAGGGCCCAAGCGCATCATCAAGTGTTCGAATGCGGCAAGAGGGC	120				
Db	61	CGCAGCAATTCGAAGGGCCCAAGCGCATCATCAAGTGTTCGAATGCGGCAAGAGGGC	120				
Qy	121	CACATCGCCGCAACTGCGCGCCGCCCGCAAGAGGGCTGCTGGAAGTGCAGCAAGGAG	180				
Db	121	CACATCGCCGCAACTGCGCGCCGCCCGCAAGAGGGCTGCTGGAAGTGCAGCAAGGAG	180				
Qy	181	GGCCACCATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCC	240				
Db	181	GGCCACCATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCC	240				
Qy	241	TTCCCCAGGGCAAGGCCCGCGAGTTCCCGAGCGAGAGAACCGCGCAACAGCCCCACC	300				
Db	241	TTCCCCAGGGCAAGGCCCGCGAGTTCCCGAGCGAGAGAACCGCGCAACAGCCCCACC	300				
Qy	301	AGCGCGAGCTGAGGTGCGCGCGCAACCCCGGAGCGAGGCCCGCGCGAGCGCCAG	360				
Db	301	AGCGCGAGCTGAGGTGCGCGCGCAACCCCGGAGCGAGGCCCGCGCGAGCGCCAG	360				
Qy	361	GGCACCTGAACTTCCCGCCAGATCACCTGTGCGAGCGCCCTTGTGTGAGCATCAAGGTG	420				
Db	361	GGCACCTGAACTTCCCGCCAGATCACCTGTGCGAGCGCCCTTGTGTGAGCATCAAGGTG	420				
Qy	421	GGCGGCGAGATCAAGAGGGCCCTGTGGAACAACCGCGCGCGAAGCATCTGTGAGGAG	480				
Db	421	GGCGGCGAGATCAAGAGGGCCCTGTGGAACAACCGCGCGCGAAGCATCTGTGAGGAG	480				
Qy	481	ATGAGCTTGGCCGCAAGTGGAGCCCAAGATCATCGCGGCATCGCGGCTTCATCAAG	540				
Db	481	ATGAGCTTGGCCGCAAGTGGAGCCCAAGATCATCGCGGCATCGCGGCTTCATCAAG	540				
Qy	541	GTGCGCCAGTACGACCATGCTGATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTG	600				
Db	541	GTGCGCCAGTACGACCATGCTGATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTG	600				
Qy	601	CTGATCGGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGAGCCAGCTGGGCTGC	660				
Db	601	CTGATCGGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGAGCCAGCTGGGCTGC	660				
Qy	661	ACCTGAACTTCCCATTCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG	720				
Db	661	ACCTGAACTTCCCATTCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG	720				
Qy	721	GAGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780				
Db	721	GACGCCCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780				
Qy	781	ATCTGAGAGAGATGAGAGAGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTAC	840				
Db	781	ATCTGAGAGAGATGAGAGAGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTAC	840				
Qy	841	AACACCCCGGTGTTCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	900				
Db	841	AACACCCCGGTGTTCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	900				
Qy	901	TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTTGGGAGGTGAGCTGGGCGATCCCCAC	960				
Db	901	TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTTGGGAGGTGAGCTGGGCGATCCCCAC	960				
Qy	961	CCCGCGGCTGAAG	1020				
Db	961	CCCGCGGCTGAAG	1020				

Qy	1021	AGCGTCCCTTGACGAGGACTTCCGCAAGTACACCGCCTTCAACCATCCCCAGCATCAAC	1080				
Db	1021	AGCGTCCCTTGACGAGGACTTCCGCAAGTACACCGCCTTCAACCATCCCCAGCATCAAC	1080				
Qy	1081	AACGAGACCCCGGATCCGCTACCAAGTACAAGCTGCTGCCAGGGCTGGGAAGGGGAGC	1140				
Db	1081	AACGAGACCCCGGATCCGCTACCAAGTACAAGCTGCTGCCAGGGCTGGGAAGGGGAGC	1140				
Qy	1141	CCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGGCCCAACCC	1200				
Db	1141	CCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGGCCCAACCC	1200				
Qy	1201	GAGATCGTATCTTACCAAGTACATGAGCAGCTTACGCTGGGAGCGACCTCTGGAGTCGGC	1260				
Db	1201	GAGATCGTATCTTACCAAGTACATGAGCAGCTTACGCTGGGAGCGACCTCTGGAGTCGGC	1260				
Qy	1261	CAGCACCGGCCAAGATTCGAGGAGCTGGCGAAGCACTGTGCTGCTGGGGCTTACCAACC	1320				
Db	1255	CAGCACCGGCCAAGATTCGAGGAGCTGGCGAAGCACTGTGCTGCTGGGGCTTACCAACC	1314				
Qy	1321	CCCGACAAGACGACCAAGAGGAGCCCTTCTGTGGATGGGCTACGAGCTGCACCC	1380				
Db	1315	CCCGACAAGAGCAGCAAGAGGAGCCCTTCTGTGCCAT-----CGAGCTGCACCC	1368				
Qy	1381	GACAACTGGACCGTGCAGGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC	1440				
Db	1369	GACAACTGGACCGTGCAGGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC	1428				
Qy	1441	ATCCAGAACTGGTGGGCAAGCTGAACTGGGGCAGCGAGATCTACCCCGGCGATCAAGGTG	1500				
Db	1429	ATCCAGAACTGGTGGGCAAGCTGAACTGGGGCAGCGAGATCTACCCCGGCGATCAAGGTG	1488				
Qy	1501	CGCCAGCTGTGCAAGCTGTGCGCGGGCCCAAGGCCCTGACCCACATCTGTCGCCCTGACC	1560				
Db	1489	CGCCAGCTGTGCAAGCTGTGCGCGGGCCCAAGGCCCTGACCCACATCTGTCGCCCTGACC	1548				
Qy	1561	GAGGAGGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGACGGC	1620				
Db	1549	GAGGAGGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGACGGC	1608				
Qy	1621	GTGTACTACGACCCCAAGGACCTGTGTGGCCGAGATCCAGAAAGCAGGGCCAGACAG	1680				
Db	1609	GTGTACTACGACCCCAAGGACCTGTGTGGCCGAGATCCAGAAAGCAGGGCCAGACAG	1668				
Qy	1681	TGGACCTTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCCAAG	1740				
Db	1669	TGGACCTTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCCAAG	1728				
Qy	1741	ATCGCACCGGCCCAACCAACGACGCTGAAGCAGCTGACCGAGGCCGTGACAGAGATCGCC	1800				
Db	1729	ATCGCACCGGCCCAACCAACGACGCTGAAGCAGCTGACCGAGGCCGTGACAGAGATCGCC	1788				
Qy	1801	ATGGAGAGATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAGGAG	1860				
Db	1789	ATGGAGAGATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAGGAG	1848				
Qy	1861	ACCTGGGAGACCTGTGTGACCGGACCTACTGCGAGCGCCACCTGGATCCCGAGTGGAGTTC	1920				
Db	1849	ACCTGGGAGACCTGTGTGACCGGACCTACTGCGAGCGCCACCTGGATCCCGAGTGGAGTTC	1908				
Qy	1921	GTGAACACCCCGCTTGTGTGAAAGCTGTGGTACACAGCTGGAGAGGAGCCCATCATCGGC	1980				
Db	1909	GTGAACACCCCGCTTGTGTGAAAGCTGTGGTACACAGCTGGAGAGGAGCCCATCATCGGC	1968				
Qy	1981	GCCGAGACCTTCTACGTGAGCGGGCCCAACCGCGAGACCAAGATCGGCAAGGGCGGC	2040				
Db	1969	GCCGAGACCTTCTACGTGAGCGGGCCCAACCGCGAGACCAAGATCGGCAAGGGCGGC	2028				
Qy	2041	TACGTGACCGACCGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG	2100				
Db	2029	TACGTGACCGACCGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG	2088				

```
Qy 2101 ACCGAGTGCAGGCGCATCCAGTGGCTTGCAGGACAGCGGCGAGGAGGTGAACATCGTG 2160
Db 2089 ACCGAGTGCAGGCGCATCCAGTGGCTTGCAGGACAGCGGCGAGGAGGTGAACATCGTG 2148
Qy 2161 ACCGACAGCAGTAGTACGCCCTGGGCGATCATCCAGGCCAGCCGAGCAAGAGCGAGCGAG 2220
Db 2149 ACCGACAGCAGTAGTACGCCCTGGGCGATCATCCAGGCCAGCCGAGCAAGAGCGAGCGAG 2208
Qy 2221 CTGCTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2280
Db 2209 CTGCTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2268
Qy 2281 CCCGCCACAAAGGCGATCGGCGGCAACGAGCAGATGCACAGCTGGTGAAGAGGGCATC 2340
Db 2269 CCCGCCACAAAGGCGATCGGCGGCAACGAGCAGATGCACAGCTGGTGAAGAGGGCATC 2328
Qy 2341 CGCAAGTGTCTTCTTGGAGCGGATCGATGGCGGCGATCGTATCTACCACTATACGAC 2400
Db 2329 CGCAAGTGTCTTCTTGGAGCGGATCGATGGCGGCGATCGTATCTACCACTATACGAC 2388
Qy 2401 GACCTGTACGTGGCAGCGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
Db 2389 GACCTGTACGTGGCAGCGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2448
Qy 2461 GGTGAATTC 2469
Db 2449 GGTGAATTC 2457
```

## RESULT 8

```
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MECEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBERG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMMW_C
US-10-190-435-43
```

```
Query Match 97.3%; Score 2401.8; DB 15; Length 2445;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;
```

```
Qy 7 GCCACCATGCGCGGCGCATGAGCCAGGCCACCGCCCAACATCTCTGATGCGCGCAGC 66
Db 1 GCCACCATGCGCGGCGCATGAGCCAGGCCACCGCCCAACATCTCTGATGCGCGCAGC 60
Qy 67 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGGCCACATC 126
Db 61 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGGCCACATC 120
Qy 127 GCGCGCAACTGCGCGCGCCCCCGCAAGAAGGGCTGTGGAAGTGTGCGCAAGAGGGGCCAC 186
Db 121 GCGCGCAACTGCGCGCGCCCCCGCAAGAAGGGCTGTGGAAGTGTGCGCAAGAGGGGCCAC 180
Qy 187 CAGATGAAGGACTGACCGGAGCGCGCAGGCCAACTTCTTCGCGGAGGACCTTGGCCTTCCCC 246
Db 181 CAGATGAAGGACTGACCGGAGCGCGCAGGCCAACTTCTTCGCGGAGGACCTTGGCCTTCCCC 240
```

```
Qy 247 CAGGCGAAGGCCCGCGAGTTTCCCGAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGCGC 306
Db 241 CAGGCGAAGGCCCGCGAGTTTCCCGAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGCGC 300
Qy 307 GAGCTGCAGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGGCGGCGAGCGCCAGGGCACC 366
Db 301 GAGCTGCAGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGGCGGCGAGCGCCAGGGCACC 360
Qy 367 CTGAATCTTCCCGCAGATCACCTCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGTGTGGCGGCG 426
Db 361 CTGAATCTTCCCGCAGATCACCTCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGTGTGGCGGCG 420
Qy 427 CAGATCAAGAGGCGCTTGTGGAACAACCGGCGCGCAACACACCTGTGTGGAAGAGATGAGC 486
Db 421 CAGATCAAGAGGCGCTTGTGGAACAACCGGCGCGCAACACACCTGTGTGGAAGAGATGAGC 480
Qy 487 CTGCCCCGCAAGTGAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTTCATCAAGGTGGCG 546
Db 481 CTGCCCCGCAAGTGAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTTCATCAAGGTGGCG 540
Qy 547 CAGTACGACCAAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATC 606
Db 541 CAGTACGACCAAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATC 600
Qy 607 GGCCCCACCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCGTG 666
Db 601 GGCCCCACCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCGTG 660
Qy 667 AACTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATGAGCGGC 726
Db 661 AACTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATGAGCGGC 720
Qy 727 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 786
Db 721 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 780
Qy 787 GAGGAGATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCGGAGAAACCCCTACAACACC 846
Db 781 GAGGAGATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCGGAGAAACCCCTACAACACC 840
Qy 847 CCGTGTTCGCCATCAAGAAGAGGACAGACCAAGTGGCGGCAAGCTGGTGGACTTCGC 906
Db 841 CCGTGTTCGCCATCAAGAAGAGGACAGACCAAGTGGCGGCAAGCTGGTGGACTTCGC 900
Qy 907 GAGCTGAACAAGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCATCCCCACCCCGCC 966
Db 901 GAGCTGAACAAGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCATCCCCACCCCGCC 960
Qy 967 GGCTTGAAGAAGAAAGAGCGTGACCGTGTGACGTGGCGGCGACGCTTACTTCAGCGTG 1026
Db 961 GGCTTGAAGAAGAAAGAGCGTGACCGTGTGACCGTGGCGGCGACGCTTACTTCAGCGTG 1020
Qy 1027 CCCTTGAAGAGGACTTCCGCAAGTACACCGCTTACCATCCCGCAGCATCAACAGAG 1086
Db 1021 CCCTTGAAGAGGACTTCCGCAAGTACACCGCTTACCATCCCGCAGCATCAACAGAG 1080
Qy 1087 ACCCCGCGCATCCGCTACCAAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCCAGC 1146
Db 1081 ACCCCGCGCATCCGCTACCAAGTACAACGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCCAGC 1140
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTTCGCGCGCCCGCAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTTCGCGCGCCCGCAACCCCGAGATC 1200
Qy 1207 GTGATCTTACAGTACATGGAACCTGTACGTGGGCGAGCACTTGGAGATCGGCGCAGAC 1266
Db 1201 GTGATCTTACCA-----GGCCCCCTTGTAGCTGGGCGAGCACTTGGAGATCGGCGCAGAC 1254
Qy 1267 CCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTTCACACCCCCGAC 1326
Db 1255 CCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTTCACACCCCCGAC 1314
```



QY 1327 AAGAAGCACAGAGAGAGCCCTTCTCTGTGATGGGCTACGAGCTGCAACCCCGACAAG 1386  
 DB 1315 AAGAAGCACAGAGAGAGCCCTTCTCTGTGATGGGCTACGAGCTGCAACCCCGACAAG 1368  
 QY 1387 TGGACCGTGCAGCCATCGAGCTGCGCGGAGAGAGAGCTGGACCGTGAAACGATCCAG 1446  
 DB 1369 TGGACCGTGCAGCCATCGAGCTGCGCGGAGAGAGAGCTGGACCGTGAAACGATCCAG 1428  
 QY 1447 AAGCTGTTGGGCAAGCTGAACCTGGGCGGAGAGAGCTGACCCCGGATCAAGGTGCGCCAG 1506  
 DB 1429 AAGCTGTTGGGCAAGCTGAACCTGGGCGGAGAGAGCTGACCCCGGATCAAGGTGCGCCAG 1488  
 QY 1507 CTGTGAAGCTGTGCGCGGCGGCAAGGCCCTTCAAGACATCTGCTGCGCCCTGACCGAGGAG 1566  
 DB 1489 CTGTGAAGCTGTGCGCGGCGGCAAGGCCCTTCAAGACATCTGCTGCGCCCTGACCGAGGAG 1548  
 QY 1567 GCCGAGCTGAGCTGGCGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGCAACGGCGGTGAC 1626  
 DB 1549 GCCGAGCTGAGCTGGCGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGCAACGGCGGTGAC 1608  
 QY 1627 TACGACCCAGCAAGGACCTGTGGCGGAGATCCAGAGCAGGCGGCAACGACAGTGGACC 1686  
 DB 1609 TACGACCCAGCAAGGACCTGTGGCGGAGATCCAGAGCAGGCGGCAACGACAGTGGACC 1668  
 QY 1687 TACCAGATCTACAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCAAGATGCGC 1746  
 DB 1669 TACCAGATCTACAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCAAGATGCGC 1728  
 QY 1747 ACCGCCACACCAACGAGCTGAAGAGCTGACCGAGCGCGTGAGAGATCGCCATGGAG 1806  
 DB 1729 ACCGCCACACCAACGAGCTGAAGAGCTGACCGAGCGCGTGAGAGATCGCCATGGAG 1788  
 QY 1807 AGCATCGTGTCTGGGCGAGAGCCCGCAAGTTCGCTGCGCCATCCAGAGGAGACTGG 1866  
 DB 1789 AGCATCGTGTCTGGGCGAGAGCCCGCAAGTTCGCTGCGCCATCCAGAGGAGACTGG 1848  
 QY 1867 GAGACCTGTGTGACCGACTACTGTGCGAGCCACCTGGATCCCGAGTGGGAGTTCTGTAAC 1926  
 DB 1849 GAGACCTGTGTGACCGACTACTGTGCGAGCCACCTGGATCCCGAGTGGGAGTTCTGTAAC 1908  
 QY 1927 ACCGCCCTCTGTGTAAGCTGTGGTACAGCTGGAGAGAGCCCATCATCTGCGCGCGAG 1986  
 DB 1909 ACCGCCCTCTGTGTAAGCTGTGGTACAGCTGGAGAGAGCCCATCATCTGCGCGCGAG 1968  
 QY 1987 ACCTTCTACGTGACGCGCGCGGCAACCGGAGACCAAGATTCGCGAGGAGCGGCTAGCTG 2046  
 DB 1969 ACCTTCTACGTGACGCGCGCGGCAACCGGAGACCAAGATTCGCGAGGAGCGGCTAGCTG 2028  
 QY 2047 ACCGACCGGGCGCGCAGAGATCGTGAGCTTGACCGAGACCAACCAAGAGACCGAG 2106  
 DB 2029 ACCGACCGGGCGCGCAGAGATCGTGAGCTTGACCGAGACCAACCAAGAGACCGAG 2088  
 QY 2107 CTGAGGCCATTCAGCTGGCCCTGCGAGGACAGCGGAGCGAGGTGAACATCTGTGACCGAC 2166  
 DB 2089 CTGAGGCCATTCAGCTGGCCCTGCGAGGACAGCGGAGCGAGGTGAACATCTGTGACCGAC 2148  
 QY 2167 AGCAGTACCGCTGGGCAATCTACCGGCGGAGCGGCGGAGAGCGAGCGGCTGGTG 2226  
 DB 2149 AGCAGTACCGCTGGGCAATCTACCGGCGGAGCGGCGGAGAGCGGCTGGTG 2208  
 QY 2227 AACAGATCATCGAGAGCTGATCAAGAGGAGAGGTGATCTGAGCTGGGTCGCGCC 2286  
 DB 2209 AACAGATCATCGAGAGCTGATCAAGAGGAGAGGTGATCTGAGCTGGGTCGCGCC 2268  
 QY 2287 CAAGAAGGATCGGCGGCAACGAGAGATCGAACAGCTGTGTGAGAGGCGCATCGCAAG 2346  
 DB 2269 CAAGAAGGATCGGCGGCAACGAGAGATCGAACAGCTGTGTGAGAGGCGCATCGCAAG 2328  
 QY 2347 GTCTGTTCCTGAGCGGATCGATGCGGCGATCGTGTATCTACAGTACATGAGACGCTG 2406  
 DB 2329 GTCTGTTCCTGAGCGGATCGATGCGGCGATCGTGTATCTACAGTACATGAGACGCTG 2388  
 QY 2407 TACGTGGGAGCGGCGGCGCTTAGGATCGATTAAAGGCTTCCCGGGGCTAGCACCGGT 2463

DB 2389 TACGTGGGAGCGGCGGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2445  
 RESULT 9  
 US-10-190-305A-37  
 ; Sequence 37, Application US/10190305A  
 ; Publication No. US20030198621A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEDEDE, Jan  
 ; APPLICANT: BARNETT, Susan  
 ; APPLICANT: LIAN, Ying  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
 ; FILE REFERENCE: 2302-18702 / 18702.002  
 ; CURRENT APPLICATION NUMBER: US/10/190,305A  
 ; CURRENT FILING DATE: 2002-07-05  
 ; NUMBER OF SEQ ID NOS: 93  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 37  
 ; LENGTH: 2445  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:  
 ; OTHER INFORMATION: p2Pol.opt.YMMW\_C  
 US-10-190-305A-37  
 Query Match 97.3%; Score 2401.8; DB 16; Length 2445;  
 Best Local Similarity 99.2%; Pred. No. 0;  
 Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;  
 QY 7 GCCACCATCGCGAGGCCATGAGCCAGGCGCACCGAGCGCAACATCTCTGATGACGCGCAGC 66  
 DB 1 GCCACCATCGCGAGGCCATGAGCCAGGCGCACCGAGCGCAACATCTCTGATGACGCGCAGC 60  
 QY 67 AACTTCAAGGGGCGGCGGCGCATCATCAAGTGTCTTCAACTGCGGCGAAGGAGGCGCCATC 126  
 DB 61 AACTTCAAGGGGCGGCGGCGCATCATCAAGTGTCTTCAACTGCGGCGAAGGAGGCGCCATC 120  
 QY 127 GCCCGCAACTGCGCGCGCGCGCGCGAGAGGGCTGTGGAAGTGTGCGGCAAGGAGGCGCAC 186  
 DB 121 GCCCGCAACTGCGCGCGCGCGCGCGAGAGGGCTGTGGAAGTGTGCGGCAAGGAGGCGCAC 180  
 QY 187 CAGATCAAGGAGCTGACCGAGCGCGCGCGCAACTCTTCTCGCGAGGACCTGGGCTTCCCC 246  
 DB 181 CAGATCAAGGAGCTGACCGAGCGCGCGCGCAACTCTTCTCGCGAGGACCTGGGCTTCCCC 240  
 QY 247 CAGGGCAAGGCGCGCGAGTTCCCGAGGAGAGAACCGCGCGCAACAGCGCCACAGCGCGC 306  
 DB 241 CAGGGCAAGGCGCGCGAGTTCCCGAGGAGAGAACCGCGCGCAACAGCGCCACAGCGCGC 300  
 QY 307 GAGCTGAGGTGCGCGCGGCGCAACCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGC 366  
 DB 301 GAGCTGAGGTGCGCGCGGCGCAACCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
 QY 367 CTGAACCTTCCCGAGATCACTCTGTGCGAGCGCGCGCTGTGAGCATCAAGGTGGGCGGC 426  
 DB 361 CTGAACCTTCCCGAGATCACTCTGTGCGAGCGCGCGCTGTGAGCATCAAGGTGGGCGGC 420  
 QY 427 CAGATCAAGGAGGCGCGCTGTGGAACCGCGCGCGCGAGCACCGGTGTGGAAGATGAGC 486  
 DB 421 CAGATCAAGGAGGCGCGCTGTGGAACCGCGCGCGCGAGCACCGGTGTGGAAGATGAGC 480  
 QY 487 CTGCCCGGCAAGTGGAGGCCAGATGATGCGCGGGCATCGCGGGCTTTCATCAAGGTGCGC 546  
 DB 481 CTGCCCGGCAAGTGGAGGCCAGATGATGCGCGGGCATCGCGGGCTTTCATCAAGGTGCGC 540  
 QY 547 CAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAAGGCCCATCGGCAACCGTGTGATC 606  
 DB 541 CAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAAGGCCCATCGGCAACCGTGTGATC 600  
 QY 607 GGCCCCACCCCGTGTGAACATCATCGCGCGCAACATCTGTACCCAGCTGTGGGCTGACCCCTG 666

Db 601 GGGCCCCCCCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660  
Qy 667 AACTTCCCCCATCAGCCCCCATCGAGACCGTGCCTGAGCTGAAAGCCCGGCGCATCGAAGGC 726  
Db 661 AACTTCCCCCATCAGCCCCCATCGAGACCGTGCCTGAGCTGAAAGCCCGGCGCATCGAAGGC 720  
Qy 727 CCCAAGGTGAAGAGTGGCCCCCTGACGAGGAGAGATCAAGGGCCCTGACCGCCATCTGC 786  
Db 721 CCCAAGGTGAAGAGTGGCCCCCTGACGAGGAGAGATCAAGGGCCCTGACCGCCATCTGC 780  
Qy 787 GAGCAGATGGAGAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGAGACCCCTTCAACACC 846  
Db 781 GAGGAGATGGAGAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGAGACCCCTTCAACACC 840  
Qy 847 CCCGTGTTCCGCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906  
Db 841 CCCGTGTTCCGCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900  
Qy 907 GAGCTGAACAGCGACCCAGGACTTCTGGAGGTGAGCTGGGGATCCGCCACCCCGCC 966  
Db 901 GAGCTGAACAGCGACCCAGGACTTCTGGAGGTGAGCTGGGGATCCGCCACCCCGCC 960  
Qy 967 GGCCTGAAGAGAGAGAGCGTGACCGTGTGAGCTGGCGAGCGCTACTTTCAGCGTG 1026  
Db 961 GGCCTGAAGAGAGAGAGCGTGACCGTGTGAGCTGGCGAGCGCTACTTTCAGCGTG 1020  
Qy 1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAACAAAGAG 1086  
Db 1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAACAAAGAG 1080  
Qy 1087 ACCCCCGGCAATCCGCTACAGTAAACAGTGTGCTCCCGAGGCTGGAAGGGCAGCCCCAGC 1146  
Db 1081 ACCCCCGGCAATCCGCTACAGTAAACAGTGTGCTCCCGAGGCTGGAAGGGCAGCCCCAGC 1140  
Qy 1147 ATCTTCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACAGTACATGAGAGCTGTACGTGGGAGAGGACCTGGAGATCGGCGAGCAC 1266  
Db 1201 GTGATCTACCA-----GGCCCCCTGTACGTGGGAGAGGACCTGGAGATCGGCGAGCAC 1254  
Qy 1267 CGGCCAAGATCGAGGAGCTGCGAAGCAGCTGTGCTGGGCTTTCACACACCCCGAC 1326  
Db 1255 CGGCCAAGATCGAGGAGCTGCGAAGCAGCTGTGCTGGGCTTTCACACACCCCGAC 1314  
Qy 1327 AAGAGCACCAAGAGAGCCCCCTTCTGTGTGAGTGGGCTACGAGCTGCACCCCGACAAG 1386  
Db 1315 AAGAGCACCAAGAGAGCCCCCTTCTGTGTGAGTGGGCTACGAGCTGCACCCCGACAAG 1368  
Qy 1387 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGGTGAACGACATCCAG 1446  
Db 1369 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGGTGAACGACATCCAG 1428  
Qy 1447 AAGCTGTGGGCAAGCTGAATCGGGCCAGCAGATCTACCCCGGCGATCAAGGTGCGCCAG 1506  
Db 1429 AAGCTGTGGGCAAGCTGAATCGGGCCAGCAGATCTACCCCGGCGATCAAGGTGCGCCAG 1488  
Qy 1507 CTGTGCAAGCTGTGCGCGGGCGCAAGCCCTGACCGACATCTGTGCCCTTGCACGAGAG 1566  
Db 1489 CTGTGCAAGCTGTGCGCGGGCGCAAGCCCTGACCGACATCTGTGCCCTTGCACGAGAG 1548  
Qy 1567 GCCGAGCTGGAGCTGGCGCGAGAACCGGAGATCTGTGGCGAGCCCGTGACGGGCTGTAC 1626  
Db 1549 GCCGAGCTGGAGCTGGCGCGAGAACCGGAGATCTGTGGCGAGCCCGTGACGGGCTGTAC 1608  
Qy 1627 TACGACCCCGAGAGGACCTGGTGGCGCGAGATCCAGAGCAGGGCCACGACCAAGTGGACC 1686  
Db 1609 TACGACCCCGAGAGGACCTGGTGGCGCGAGATCCAGAGCAGGGCCACGACCAAGTGGACC 1668  
Qy 1687 TACGAGATCTACAGGAGCCCTTCAAGAACTGTAAGAACCGGCAAGTACGCGCAAGTACGCG 1746

Db 1669 TACGAGATCTACAGGAGCCCTTCAAGAACTGTAAGACCGGCAAGTACGCGCAAGTACGCG 1728  
Qy 1747 ACCGCCCCACACCAACGAGCTGAAGAGCAGCTGACCGAGCGCGTGCAAGAGATCGCCATGGAG 1806  
Db 1729 ACCGCCCCACACCAACGAGCTGAAGAGCAGCTGACCGAGGCGGTGCAAGAGATCGCCATGGAG 1788  
Qy 1807 AGCATCTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGAGACCTGG 1866  
Db 1789 AGCATCTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGAGACCTGG 1848  
Qy 1867 GAGACCTGTGTGAGCCGACTACTGTGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1926  
Db 1849 GAGACCTGTGTGAGCCGACTACTGTGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1908  
Qy 1927 ACCGCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGAGAGCCCATCATCGGGCCCGAG 1986  
Db 1909 ACCGCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGAGAGCCCATCATCGGGCCCGAG 1968  
Qy 1987 ACCTTCTACGTGTGAGCGGCGGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGGTACGTG 2046  
Db 1969 ACCTTCTACGTGTGAGCGGCGGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGGTACGTG 2028  
Qy 2047 ACCGACCGGGCGCGCGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2106  
Db 2029 ACCGACCGGGCGCGCGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2088  
Qy 2107 CTGCAGGCCCATCCAGCTGGGCCCTGTCAGGACAGCGGCGAGCGGTGAACATCGTGAACGAC 2166  
Db 2089 CTGCAGGCCCATCCAGCTGGGCCCTGTCAGGACAGCGGCGAGCGGTGAACATCGTGAACGAC 2148  
Qy 2167 AGCAGTACGCTTGGGCGATCATCCAGGCCAGCCCCGACCAAGAGCGAGCGAGCTGTG 2226  
Db 2149 AGCAGTACGCTTGGGCGATCATCCAGGCCAGCCCCGACCAAGAGCGAGCGAGCTGTG 2208  
Qy 2227 AACAGATCATCGAGCAGCTGTCAAGAGGAGAGGTGTACCTGAGCTGGTGGTCCCGCC 2286  
Db 2209 AACAGATCATCGAGCAGCTGTCAAGAGGAGAGGTGTACCTGAGCTGGTGGTCCCGCC 2268  
Qy 2287 CACAAGGGCATCGGCGCAACGAGCAGATGCACAACTGGTGAGCAAGGGGATCCGCAAG 2346  
Db 2269 CACAAGGGCATCGGCGCAACGAGCAGATGCACAACTGGTGAGCAAGGGGATCCGCAAG 2328  
Qy 2347 GTGCTGTCTTGGACGCGCATCGATGGCGGCGATCGTGTACCAAGTACATGAGACGCTG 2406  
Db 2329 GTGCTGTCTTGGACGCGCATCGATGGCGGCGATCGTGTACCAAGTACATGAGACGCTG 2388  
Qy 2407 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2463  
Db 2389 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2445

## RESULT 10

US-10-435-9

; Sequence 9, Application US/10190435

; Publication No. US20030143249A1

; GENERAL INFORMATION:

; APPLICANT: ZUR MEGEDE, Jan

; APPLICANT: BARNETT, Susan W.

; APPLICANT: LIAN, Ying

; APPLICANT: ENGELBRECHT, Susan

; APPLICANT: VAN RENSBURG, Estrelita J.

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C

; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

; FILE REFERENCE: PP18133.003 / 2302-18133

; CURRENT APPLICATION NUMBER: US/10190,435

; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 3930

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmut_C US-10-190-435-9									
Query Match		97.0%	Score 2394.8;	DB 15;	Length 3930;				
Best Local Similarity		99.2%	Pred. No. 0;						
Matches 2431;		Conservative	0;	Mismatches	7;	Indels	12;	Gaps	2;
QY	14	TGCGCGAGGCCATGAGCGAGCGCCACAGCGCCCAACATCTCTGATGCGAGCGCAGCAACTTCA	73						
DB	1487	TCGCCAGGCCATGAGCGAGCGCCACAGCGCCCAACATCTCTGATGCGAGCGCAGCAACTTCA	1546						
QY	74	AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTTGGCGCAAGGAGGGGCCATTCGCCCGCA	133						
DB	1547	AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTTGGCGCAAGGAGGGGCCATTCGCCCGCA	1606						
QY	134	ACTGCGCGCCCCCGCAAGAGGGCTGTCTGGAAGTGCGGCAAGGAGGGGCCACAGATGA	193						
DB	1607	ACTGCGCGCCCCCGCAAGAGGGCTGTCTGGAAGTGCGGCAAGGAGGGGCCACAGATGA	1666						
QY	194	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCTTCCCGCAGGGCA	253						
DB	1667	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCTTCCCGCAGGGCA	1726						
QY	254	AGGCCCGGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGCCCAACAGCCGCGAGCTGC	313						
DB	1727	AGGCCCGGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGCCCAACAGCCGCGAGCTGC	1786						
QY	314	AGGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	373						
DB	1787	AGGTGCGCGGCGAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1846						
QY	374	TCCCCCAGATCACCTCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGTGGGGCGGCGCAGATCA	433						
DB	1847	TCCCCCAGATCACCTCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGTGGGGCGGCGCAGATCA	1906						
QY	434	AGGAGGCCCTGTGTGACACCGCGCGCGCAGACACCGTGTCTGTGAGGAGATGAGCTTGCCCG	493						
DB	1907	AGGAGGCCCTGTGTGACACCGCGCGCGCAGACACCGTGTCTGTGAGGAGATGAGCTTGCCCG	1966						
QY	494	GCAAGTGAAGCCCAAGATGATCGGGCGCATTCGGCGGCTTTCATCAAGGTGCGCGCAGTACG	553						
DB	1967	GCAAGTGAAGCCCAAGATGATCGGGCGCATTCGGCGGCTTTCATCAAGGTGCGCGCAGTACG	2026						
QY	554	ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCA	613						
DB	2027	ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCA	2086						
QY	614	CCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCC	673						
DB	2087	CCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCC	2146						
QY	674	CCATCAGCCCATCGAGACCGTGCCTGTGAGCTGAAGCCCGGCATGGAAGCGGCCCAAGG	733						
DB	2147	CCATCAGCCCATCGAGACCGTGCCTGTGAGCTGAAGCCCGGCATGGAAGCGGCCCAAGG	2206						
QY	734	TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGCGAGGAGA	793						
DB	2207	TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGCGAGGAGA	2266						
QY	794	TGGAGAAGGAGGCAAGATCACAAGATCGGCCCGCGAGAACCCCTACAAACACCCCGTGT	853						
DB	2267	TGGAGAAGGAGGCAAGATCACAAGATCGGCCCGCGAGAACCCCTACAAACACCCCGTGT	2326						
QY	854	TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAGCTGGTGGACTTCCGCGAGCTGA	913						
DB	2327	TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAGCTGGTGGACTTCCGCGAGCTGA	2386						
QY	914	ACAAGCGCACCCAGGACTTCTTGGAGGTGAGCTGGGCATCCCGCAGCCCGCGCGCTGA	973						
DB	2387	ACAAGCGCACCCAGGACTTCTTGGAGGTGAGCTGGGCATCCCGCAGCCCGCGCGCTGA	2446						
QY	974	AGAAGAAGAGAGCGTGACCGTGTGSAACGTGGCGCAGCGCTTACTTCAGCGTGCCTTGG	1033						

2447	AGAAGAAGAAGAGCGTGACCGTGTCTGACGTGGCGACCGCTTACTTTACGCGTCCCGCTGG	2506
1034	ACGAGGACTTTCCGCAAGTATACCGCCTTCCACCATCCCGAGCATCAACAACAGACACCCCG	1093
2507	ACGAGGACTTTCCGCAAGTATACCGCCTTCCACCATCCCGAGCATCAACAACAGACACCCCG	2566
1094	GCATCCGCTACCAAGTACAAAGTGTCTGCCCAAGGGCTTGAAGGCGACCCCGAGCATTTCC	1153
2567	GCATCCGCTACCAAGTATACAAAGTGTCTGCCCAAGGGCTTGAAGGCGACCCCGAGCATTTCC	2626
1154	AGAGCAGCATGACCAAGATCTTGAGGCTTCCGCGCCCGCAACCCCGAGATCGTGATCT	1213
2627	AGAGCAGCATGACCAAGATCTTGAGGCTTCCGCGCCCGCAACCCCGAGATCGTGATCT	2686
1214	ACCAGTACATGACAGCACTGTAGCTGGGCGAGGACCTTGAGATCGGCGCAGCACCGCGCCA	1273
2687	ACCA-----GGCCCCCTGTAGCTGGGCGAGGACCTTGAGATCGGCGCAGCACCGCGCCA	2740
1274	AGATCGAGGAGCTTGGCAAGCACCTTCTGCTGGGGCTTTCACCAACCCCGCAACAAGAAGC	1333
2741	AGATCGAGGAGCTTGGCAAGCACCTTCTGCTGGGGCTTTCACCAACCCCGCAACAAGAAGC	2800
1334	ACCAGAAGAGACCCCTTCTGCTGGGCTTACGAGCTGCACCCCGCAACAAGTGAACCG	1393
2801	ACCAGAAGAGACCCCTTCTGCTGGGCTTTC-----CGAGCTGCACCCCGCAACAAGTGAACCG	2854
1394	TGCAGCCCATCGAGCTTGGCGGAGAGGAGCTGGACCGTGAACGACATCCAGAAGCTGG	1453
2855	TGCAGCCCATCGAGCTTGGCGGAGAGGAGCTGGACCGTGAACGACATCCAGAAGCTGG	2914
1454	TGGGCAAGCTGAATCGGGCGAGCCAGATCTAACCCGGCATCAAGGTGCGCGAGCTGTGCA	1513
2915	TGGGCAAGCTGAATCGGGCGAGCCAGATCTAACCCGGCATCAAGGTGCGCGAGCTGTGCA	2974
1514	AGCTGCTGCGCGCGCAAGGCGCTTACCGCATCTGTGCGCCCTGACCGAGGAGCGCGAGC	1573
2975	AGCTGCTGCGCGCGCAAGGCGCTTACCGCATCTGTGCGCCCTGACCGAGGAGCGCGAGC	3034
1574	TGGAGCTGCGCGCGCAAGGCGCTTACCGCATCTGTGCGCCCTGACCGAGGAGCGCGAGC	1633
3035	TGGAGCTGCGCGCGCAAGGCGCTTACCGCATCTGTGCGCGCCCTGACCGAGGAGCGCGAGC	3094
1634	CCAGCAAGGACCTTGTGGCGCGAGATCCAGAAGCAGGCGCCACGACCAAGTGGACCTTACGAG	1693
3095	CCAGCAAGGACCTTGTGGCGCGAGATCCAGAAGCAGGCGCCACGACCAAGTGGACCTTACGAG	3154
1694	TCTACAGAGACCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGCGACCGCC	1753
3155	TCTACAGAGACCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGCGACCGCC	3214
1754	ACACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAAGATCGCCATCGGAGAGCATCG	1813
3215	ACACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAAGATCGCCATCGGAGAGCATCG	3274
1814	TGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAAGGAGACCTGGGAGACCT	1873
3275	TGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAAGGAGACCTGGGAGACCT	3334
1874	GSTGAGACCATCTGCGAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCCC	1933
3335	GSTGAGACCATCTGCGAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCCC	3394
1934	CCCTGTGAGGCTGTGTTACAGTGAAGGAGCCCATCATCTGGCGCGCGAGACCTTCT	1993
3395	CCCTGTGAGGCTGTGTTACAGTGAAGGAGCCCATCATCTGGCGCGCGAGACCTTCT	3454
1994	ACGTGGAACGCGCGCCCAACCCGCGAGACCAAGATCGGCAAGGCGGCTACGTACCGGAC	2053
3455	ACGTGGAACGCGCGCCCAACCCGCGAGACCAAGATCGGCAAGGCGGCTACGTACCGGAC	3514
2054	GGGGCGGCAAGAGATCTGTGAGCTTACCGGAGACCAACCAACAGAGACCGAGCTCGAGG	2113
3515	GGGGCGGCAAGAGATCTGTGAGCTTACCGGAGACCAACCAACAGAGACCGAGCTCGAGG	3574



1394 TGCAAGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453  
Db  
2855 TGCAAGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914  
Qy 1454 TGGGCAAGCTGAATCTGGGCGAGGACAGATCTTACCCCGGCATCAAGGTGCGCAGCTGTGCA 1513  
Db 2915 TGGGCAAGCTGAATCTGGGCGAGGACAGATCTTACCCCGGCATCAAGGTGCGCAGCTGTGCA 2974  
Qy 1514 AGCTGCTGCGCGGCGCAAGGCGCTGACCGACATCTGTGCCCTTGAACGAGGAGCGGAGC 1573  
Db 2975 AGCTGCTGCGCGGCGCAAGGCGCTGACCGACATCTGTGCCCTTGAACGAGGAGCGGAGC 3034  
Qy 1574 TGAGCTTGGCGAGAACCGGAGATCTTGGCGAGCCGCTGACCGGCTGTACTACGACC 1633  
Db 3035 TGAGCTTGGCGAGAACCGGAGATCTTGGCGAGCCGCTGACCGGCTGTACTACGACC 3094  
Qy 1634 CCAGCAAGGACCTGGTGGCGAGATCCAGAAGGAGGCGCCACGACGATGGACCTACGAGA 1693  
Db 3095 CCAGCAAGGACCTGGTGGCGAGATCCAGAAGGAGGCGCCACGACGATGGACCTACGAGA 3154  
Qy 1694 TCTACAGAGGCCCTTCAAGAACTGAGACCGGCAAGTACGCCAAGATGCGGACCGGCC 1753  
Db 3155 TCTACAGAGGCCCTTCAAGAACTGAGACCGGCAAGTACGCCAAGATGCGGACCGGCC 3214  
Qy 1754 ACACCAAGCAGCTGAAGCAGCTGACCGAGGCGCTGCAAGAGATGCCATGGAGGATCG 1813  
Db 3215 ACACCAAGCAGCTGAAGCAGCTGACCGAGGCGCTGCAAGAGATGCCATGGAGGATCG 3274  
Qy 1814 TGATCTGGGCAAGACCCCCAAGTTTCGCTGCGCCATCCAGAAGGAGACCTGGGAGACCT 1873  
Db 3275 TGATCTGGGCAAGACCCCCAAGTTTCGCTGCGCCATCCAGAAGGAGACCTGGGAGACCT 3334  
Qy 1874 GGTGACCCGACTACTGGCAGGCCACTTGATTCGCCGAGTGGGAGTTCTGTGAACACCCCCC 1933  
Db 3335 GGTGACCCGACTACTGGCAGGCCACTTGATTCGCCGAGTGGGAGTTCTGTGAACACCCCCC 3394  
Qy 1934 CCCTGGTGAAGCTGTGTATACCTGAGAGAGAGGCCCATCATCGGCGCGGAGACCTTCT 1993  
Db 3395 CCCTGGTGAAGCTGTGTATACCTGAGAGAGAGGCCCATCATCGGCGCGGAGACCTTCT 3454  
Qy 1994 ACCTGGAACGGCGCGCCAAACCGGAGACCAAGATTCGGCAAGGCGGCTACGTACCGACC 2053  
Db 3455 ACCTGGAACGGCGCGCCAAACCGGAGACCAAGATTCGGCAAGGCGGCTACGTACCGACC 3514  
Qy 2054 GGGGCGGCAAGAGATCGTGAAGCTGACCGAGACCAACCAACCAAGAGACCGAGCTGACG 2113  
Db 3515 GGGGCGGCAAGAGATCGTGAAGCTGACCGAGACCAACCAACCAAGAGACCGAGCTGACG 3574  
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGAGCGAGGTGAACATCGTGACCGACGACGAGT 2173  
Db 3575 CCATCCAGCTGGCCCTGCAGGACAGCGGAGCGAGGTGAACATCGTGACCGACGACGAGT 3634  
Qy 2174 ACGCCCTGGGCATCATCCAGGCCAGCCGACAGAGCGAGAGCTGGTGAACCGAGA 2233  
Db 3635 ACGCCCTGGGCATCATCCAGGCCAGCCGACAGAGCGAGAGCTGGTGAACCGAGA 3694  
Qy 2234 TCATCGAGAGCTGATCAAGAGAGAGAGGTGTACTGTAGCTGGTGGTGGCCGCCCAACAG 2293  
Db 3695 TCATCGAGAGCTGATCAAGAGAGAGAGGTGTACTGTAGCTGGTGGTGGCCGCCCAACAG 3754  
Qy 2294 GCATCGGCGCAACGAGCAGATCGAACGCTGGTGAAGAGGCGATCCGCAAGGTGCTGT 2353  
Db 3755 GCATCGGCGCAACGAGCAGATCGAACGCTGGTGAAGAGGCGATCCGCAAGGTGCTGT 3814  
Qy 2354 TCTTGAACGCGATCGATGGCGGATCGTGAATCTACAGTACATGAGCAGCTGTACGTGG 2413  
Db 3815 TCTTGAACGCGATCGATGGCGGATCGTGAATCTACAGTACATGAGCAGCTGTACGTGG 3874  
Qy 2414 GCAGCGCGCCCTAGGATCGATTAAGCTTCCCGGGGTAGCACCGGT 2463  
Db 3875 GCAGCGCGCCCTAGGATCGATTAAGCTTCCCGGGGTAGCACCGGT 3924

RESULT 12  
US-10-190-435-11  
; Sequence 11, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 3930  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna\_C  
US-10-190-435-11  
Query Match 96.9%; Score 2393.2; DB 15; Length 3930;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;  
Qy 14 TGGCCGAGGCCATGAGCCAGGCCACCCAGCGCCAAACATCTCTGATGCGAGCGCAACTTCA 73  
Db 1487 TCGCCGAGGCCATGAGCCAGGCCACCCAGCGCCAAACATCTCTGATGCGAGCGCAACTTCA 1546  
Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGGAGGCCACATCGCCCGCA 133  
Db 1547 AGGGCCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGGAGGCCACATCGCCCGCA 1606  
Qy 134 ACTGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGGAGGGCCACAGATGA 193  
Db 1607 ACTGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCAGCAAGGAGGGCCACAGATGA 1666  
Qy 194 AGGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCGCAGGGCA 253  
Db 1667 AGGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCGCAGGGCA 1726  
Qy 254 AGGCCCGGAGTTCCCGAGGAGCAGAACCGCGCCGACACAGCCCCACCGAGCGGAGCTGC 313  
Db 1727 AGGCCCGGAGTTCCCGAGGAGCAGAACCGCGCCGACACAGCCCCACCGAGCGGAGCTGC 1786  
Qy 314 AGGTGCGCGGACAAACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGGCACCTGAACT 373  
Db 1787 AGGTGCGCGGAGAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
Qy 374 TCCCGCCAGATCACCTGTGCGACGCGCCCTTGTGTAGCATCAAGGTGGCGCGCGAGATCA 433  
Db 1847 TCCCGCCAGATCACCTGTGCGACGCGCCCTTGTGTAGCATCAAGGTGGCGCGCGAGATCA 1906  
Qy 434 AGAGGCGCTGTGGAACAACCGCGCGCGAGACACCGCTGTGAGGAGATGAGCTGCGCCG 493  
Db 1907 AGAGGCGCTGTGTGGCACCGCGCGCGAGACACCGCTGTGAGGAGATGAGCTGCGCCG 1966  
Qy 494 GCAAGTGGAGGCCCAAGATGATCGGCGGATCGGCGGCTTTCATCAAGGTGCGCGAGTACG 553  
Db 1967 GCAAGTGGAGGCCCAAGATGATCGGCGGATCGGCGGCTTTCATCAAGGTGCGCGAGTACG 2026  
Qy 554 ACCAGATCTCTGATCGAGATCTCGGCAAGAGGCCATCGGCAACCGTGTGATCGGCCCA 613  
Db 2027 ACCAGATCTCTGATCGAGATCTCGGCAAGAGGCCATCGGCAACCGTGTGATCGGCCCA 2086  
Qy 614 CCCCCGTGAACATCATCGGCGCGCAACATCTGTGACCGAGCTGGCTGCACCCCTGAACCTTCC 673  
Db 2087 CCCCCGTGAACATCATCGGCGCGCAACATCTGTGACCGAGCTGGCTGCACCCCTGAACCTTCC 2146

```
QY 674 CCATCAGCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGCGATGAACCGGCCCAAGG 733
Db 2147 CCATCAGCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGCGATGAACCGGCCCAAGG 2206
QY 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
QY 794 TGGAGAGAGGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTTACAACACCCCCGTGT 853
Db 2267 TGGAGAGAGGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTTACAACACCCCCGTGT 2326
QY 854 TCGGCATCAAGAGAGAGAGACACCAAGTGGCGGCAAGCTGGTGAATTCGCGAGCTGA 913
Db 2327 TCGGCATCAAGAGAGAGAGACACCAAGTGGCGGCAAGCTGGTGAATTCGCGAGCTGA 2386
QY 914 ACAAGCGCACCCAGACTTCTGGGAGGTGCAAGCTGGGCAATCCCAACCCCGCGGCTGA 973
Db 2387 ACAAGCGCACCCAGACTTCTGGGAGGTGCAAGCTGGGCAATCCCAACCCCGCGGCTGA 2446
QY 974 AGAAGAGAGAGAGCGTGAACCGTGTGGACGTGGGCGAGCGCTACTTCAAGCGTGCCTCTGG 1033
Db 2447 AGAAGAGAGAGAGCGTGAACCGTGTGGACGTGGGCGAGCGCTACTTCAAGCGTGCCTCTGG 2506
QY 1034 ACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAACGAGACCCCGG 1093
Db 2507 ACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAACGAGACCCCGG 2566
QY 1094 GCATCCGCTACCACTGACAGCTGTGCTCCAGGCGTGAAGGGCAGCCCGAGCATCTTCC 1153
Db 2567 GCATCCGCTACCACTGACAGCTGTGCTCCAGGCGTGAAGGGCAGCCCGAGCATCTTCC 2626
QY 1154 AGACAGCATGACCAAGATCTCGAGGCTTCCGCGCCGCAACCCCGAGATCTGATCT 1213
Db 2627 AGACAGCATGACCAAGATCTCGAGGCTTCCGCGCCGCAACCCCGAGATCTGATCT 2686
QY 1214 ACCAGTACATGAGCGAAGCTGTGAGCGAGCAAGCTGGAGATCGGCGAGCACCGCGCA 1273
Db 2687 ACCA-----GGCCCCCTGTACGTGGCGAGCGCACTGGAGATCGGCGAGCACCGCGCA 2740
QY 1274 AGATCGAGGAGTGGCAAGCAAGCTGTGCGCTGGGGCTTACCACCCCGCGAGAGC 1333
Db 2741 AGATCGAGGAGTGGCAAGCAAGCTGTGCGCTGGGGCTTACCACCCCGCGAGAGC 2800
QY 1334 ACCAGAGGAGCCCGCTTCTGTGGATGGGCTACGAGCTGCACCCGCAAGTGGACCG 1393
Db 2801 ACCAGAGGAGCCCGCTTCTGTGGATGGGCTACGAGCTGCACCCGCAAGTGGACCG 2854
QY 1394 TGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACCGTGAAACGATCCAGAAAGCTGG 1453
Db 2855 TGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACCGTGAAACGATCCAGAAAGCTGG 2914
QY 1454 TGGCAAGCTGAACTGGGCGAGCAGATCTACCCCGGCAATCAAGTGGCGGAGCTGGCA 1513
Db 2915 TGGCAAGCTGAACTGGGCGAGCAGATCTACCCCGGCAATCAAGTGGCGGAGCTGGCA 2974
QY 1514 AGCTGTGCGCGGCGCAAGGCCCTGACCGAGATCTGCGCCCTGACCGAGGAGCGGAGC 1573
Db 2975 AGCTGTGCGCGGCGCAAGGCCCTGACCGAGATCTGCGCCCTGACCGAGGAGCGGAGC 3034
QY 1574 TGGAGCTGGCGGAGAACCGGAGATCTTGGCGAGCCCGGTGACGGGGTGTACTACGACC 1633
Db 3035 TGGAGCTGGCGGAGAACCGGAGATCTTGGCGAGCCCGGTGACGGGGTGTACTACGACC 3094
QY 1634 CCAGCAAGGAGCTGGTGGCGGAGATCCAGAGAGGGGCGAGCAGTGGACCTTACCAGA 1693
Db 3095 CCAGCAAGGAGCTGGTGGCGGAGATCCAGAGAGGGGCGAGCAGTGGACCTTACCAGA 3154
QY 1694 TCTACAGGAGCCCTTCAAGAACTGGAAGCCGCAAGTACGCCAAGATGCGCACCGCC 1753
Db 3155 TCTACAGGAGCCCTTCAAGAACTGGAAGCCGCAAGTACGCCAAGATGCGCACCGCC 3214
```

```
QY 1754 ACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCGAAGATCCCCATGGAGAGCATCG 1813
Db 3215 ACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCGAAGATCCCCATGGAGAGCATCG 3274
QY 1814 TGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTGCGCATCCAGAAGGAGACCTTGGGAGACT 1873
Db 3275 TGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTGCGCATCCAGAAGGAGACCTTGGGAGACT 3334
QY 1874 GGTGGACCGACTACTTGGCAGGCCCACTTGGATCCCGAGTGGGAGTTCTGTGAACACCCCCC 1933
Db 3335 GGTGGACCGACTACTTGGCAGGCCCACTTGGATCCCGAGTGGGAGTTCTGTGAACACCCCCC 3394
QY 1934 CCCTGGTGAAGCTGTGTACCAAGCTGGAGAGAGGCCCATCATCGGCGCCGAGACCTTCT 1993
Db 3395 CCCTGGTGAAGCTGTGTACCAAGCTGGAGAGAGGCCCATCATCGGCGCCGAGACCTTCT 3454
QY 1994 ACCTGGAGCGGCGCGCAACCCGAGAGCAAGATCGCAAGCGCGGTACGTGACCGACC 2053
Db 3455 ACCTGGAGCGGCGCGCAACCCGAGAGCAAGATCGGCAAGCGCGGTACGTGACCGACC 3514
QY 2054 GGGCCCGGCAGAAAGATCTGTAGCCTGACCGAGACCAACCAACGAGAGCCGAGCTGCAGG 2113
Db 3515 GGGCCCGGCAGAAAGATCTGTAGCCTGACCGAGACCAACCAACGAGAGCCGAGCTGCAGG 3574
QY 2114 CCATCGAGCTGGCCCTGCGAGGACAGCGGAGGAGGTGAACATCGTGACCGACCGAGT 2173
Db 3575 CCATCGAGCTGGCCCTGCGAGGACAGCGGAGGAGGTGAACATCGTGACCGACCGAGT 3634
QY 2174 ACGCCCTGGGCATCATCCAGSCCCAGCCGAGAGAGAGGAGGAGCTGGTGAACAGAG 2233
Db 3635 ACGCCCTGGGCATCATCCAGSCCCAGCCGAGAGAGAGGAGGAGCTGGTGAACAGAG 3694
QY 2234 TCATCGAGCAGCTGATCAAGAAGAGAGGTGTACTGAGCTGGGTGCCCGCCCAAGAG 2293
Db 3695 TCATCGAGCAGCTGATCAAGAAGAGAGGTGTACTGAGCTGGGTGCCCGCCCAAGAG 3754
QY 2294 GCATCGGCGGCAACGAGCAGATCGAAGCTGGTGGAGAGGGCATCCGCAAGGTGCTGT 2353
Db 3755 GCATCGGCGGCAACGAGCAGATCGAAGCTGGTGGAGAGGGCATCCGCAAGGTGCTGT 3814
QY 2354 TCCTGGAGCGCATCGATGGCGGCATCGTACCTACAGTACATGAGCAGCCTGTACGTGG 2413
Db 3815 TCCTGGAGCGCATCGATGGCGGCATCGTACCTACAGTACATGAGCAGCCTGTACGTGG 3874
QY 2414 GCAGCGCGGCCCTTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 2463
Db 3875 GCAGCGCGGCCCTTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 3924
```

## RESULT 13

```
US-10-435-58
; Sequence 58, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10190.435
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 5184
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna C
; US-10-435-58
```







```
Db 4109 TGCAGCCCATCGAGCTGCCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGG 4168
Qy 1454 TGGCAGCTGAATGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCAGCTGGCA 1513
Db 4169 TGGCAGCTGAATGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCAGCTGGCA 4228
Qy 1514 AGCTGCTGCGCGCGCCAGAGCCCTGACCGACATCGTGCCCTTGACCGAGAGCGCCGAGC 1573
Db 4229 AGCTGCTGCGCGCGCCAGAGCCCTGACCGACATCGTGCCCTTGACCGAGAGCGCCGAGC 4288
Qy 1574 TGGAGCTGGCCGAGAAACCGCGAGATCTGCGCGAGCCCGGTGCAACGGCGTGTACTAGACC 1633
Db 4289 TGGAGCTGGCCGAGAAACCGCGAGATCTGCGCGAGCCCGGTGCAACGGCGTGTACTAGACC 4348
Qy 1634 CAGCAAGAGACTGTGTGGCCGAGATCCAGAAAGCAGGSCCAGACCAAGTGAACCTACCGA 1693
Db 4349 CCAGCAAGGACCTGGTGGCCGAGATCCAGAAAGCAGGSCCAGACCAAGTGAACCTACCGA 4408
Qy 1694 TCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATCGCGACCGCCC 1753
Db 4409 TCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATCGCGACCGCCC 4468
Qy 1754 ACACCAACGACGTGAAGCAGCTGACCGAGCCCGGTGCAAGAGATCGCCATGGAGAGCATCG 1813
Db 4469 ACACCAACGACGTGAAGCAGCTGACCGAGCCCGGTGCAAGAGATCGCCATGGAGAGCATCG 4528
Qy 1814 TGATCTGGGCGAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGGAGACCTGGGAGACCT 1873
Db 4529 TGATCTGGGCGAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGGAGACCTGGGAGACCT 4588
Qy 1874 GGTGGACCGACTACTGGCAGGCACCTGGATCCCGAGTGGGAGTTCTGTAACACCCCCC 1933
Db 4589 GGTGGACCGACTACTGGCAGGCACCTGGATCCCGAGTGGGAGTTCTGTAACACCCCCC 4648
Qy 1934 CCCTGTGTGAAGCTGTGTGTAACCGTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993
Db 4649 CCCTGTGTGAAGCTGTGTGTAACCGTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCT 4708
Qy 1994 ACCTGGACCGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGGTACGTACCGACC 2053
Db 4709 ACCTGGACCGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGGTACGTACCGACC 4768
Qy 2054 GGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTGCAGG 2113
Db 4769 GGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTGCAGG 4828
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGAGCGGAGGTGAACATCGTGACCGACGACGT 2173
Db 4829 CCATCCAGCTGGCCCTGCAGGACAGCGGAGCGGAGGTGAACATCGTGACCGACGACGT 4888
Qy 2174 ACGCCCTGGGCATCATCCAGGCCAGCCCGCAAGAGCGAGAGGAGCTGGTGAACCGA 2233
Db 4889 ACGCCCTGGGCATCATCCAGGCCAGCCCGCAAGAGCGGAGAGGAGCTGGTGAACCGA 4948
Qy 2234 TCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTGGCCCGCCCAAGG 2293
Db 4949 TCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTGGCCCGCCCAAGG 5008
Qy 2294 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAAGAGGAGCATCCGCAAGGTGCTGT 2353
Db 5009 GCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAAGAGGAGCATCCGCAAGGTGCTGT 5068
Qy 2354 TCTTGAACGATCGATGGGGGATCGTGATCTACAGTATACATGAGACGACCTGTAGCTGG 2413
Db 5069 TCTTGAACGATCGATGGGGGATCGTGATCTACAGTATACATGAGACGACCTGTAGCTGG 5128
Qy 2414 GCAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db 5129 GCAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 15

US-10-190-435-13

```
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: Lian, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13
```

```
Query Match 95.7%; Score 2362.8; DB 15; Length 3531;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 2411; Conservative 0; Mismatches 27; Indels 12; Gaps 2;

Qy 14 TGGCCAGGCGCATGAGCGAGGCCACACAGCGCCAACTCTGTATGCGAGCGCAACTTCA 73
Db 1088 TGGCCAGGCGCATGAGCGAGGCCACACAGCGCGTGTATGCGAAGAGCAACTTTAAA 1147
Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCGCACATCGCCGCA 133
Db 1148 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCGCACATCGCCGCA 1207
Qy 134 ACTGCGCGCGCCCCCGCAAGAGGGGTCTGGAAGTTCGCAAGTTCGCAAGGCGCACAGATGA 193
Db 1208 ACTGCGCGCGCCCCCGCAAGAGGGGTCTGGAAGTTCGCAAGTTCGCAAGGCGCACAGATGA 1267
Qy 194 AGGACTGCAACCGAGCGCGAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCCAGGGCA 253
Db 1268 AGGACTGCAACCGAGCGCGAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCCAGGGCA 1327
Qy 254 AGSCCGCGAGTTCCCGAGCGAGCAAGACCGCGCCCAACAGCCCAACAGCGCGAGCTGC 313
Db 1328 AGSCCGCGAGTTCCCGAGCGAGCAAGACCGCGCCCAACAGCCCAACAGCGCGAGCTGC 1387
Qy 314 AGGTGCGCGCGCAAAACCCCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 373
Db 1388 AGGTGCGCGCGCAAAACCCCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1447
Qy 374 TCCCCAGATCACCTGTGGCAGCGCCCTTGTGTAGATCAAGTGGGGCGCGCAGATCA 433
Db 1448 TCCCCAGATCACCTGTGGCAGCGCCCTTGTGTAGATCAAGTGGGGCGCGCAGATCA 1507
Qy 434 AGGAGCCCTGTGGACACCGCGCGCGAGCGACACCGTGTGGAGGAGATGAGCTGCGCCG 493
Db 1508 AGGAGCCCTGTGGACACCGCGCGCGAGCGACACCGTGTGGAGGAGATGAGCTGCGCCG 1567
Qy 494 GCAAGTGGAGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAGGTGCGCGAGTACG 553
Db 1568 GCAAGTGGAGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAGGTGCGCGAGTACG 1627
Qy 554 ACAGATCTGTATCGAGATCTGGGCAAGAGGCGCATCGGCAACGCTGTATCGGCCCCA 613
Db 1628 ACAGATCTGTATCGAGATCTGGGCAAGAGGCGCATCGGCAACGCTGTATCGGCCCCA 1687
Qy 614 CCCCCGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCTGAACCTTC 673
Db 1688 CCCCCGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCTGAACCTTC 1747
Qy 674 CCATCAGCCCATCGAGACCGTGCCTGTAAGTGAAGCCCGCGCATGAGCGGCCCGCAAGG 733
Db 733
```

Db 1748 CCATCAGCCCCATCGAGACGCTGCCCGTGAAGCTGAAGCCCGGCATGGAACGGCCCCCAAGG 1807  
Qy 734 TGAAGCAGTGGCCCTGTGACCGAGAGAAAGATCAAGGCCCCCTGACCGCCATCTGCGAGAGAGA 793  
Db 1808 TGAAGCAGTGGCCCTGTGACCGAGAGAAAGATCAAGGCCCCCTGACCGCCATCTGCGAGAGAGA 1867  
Qy 794 TGGAGAGGGGGCAAGATCAACAGATCGGCGCGGAGAACCCCTACAAACCCCGCTGT 853  
Db 1868 TGGAGAGGGGGCAAGATCAACAGATCGGCGCGGAGAACCCCTACAAACCCCGCTGT 1927  
Qy 854 TCGCCATCAAGAAGAAGACAGACCAAGTGGCGCAAGCTGGTGAATTCGCGCAGCTGA 913  
Db 1928 TCGCCATCAAGAAGAAGACAGACCAAGTGGCGCAGCTGGTGAATTCGCGCAGCTGA 1987  
Qy 914 ACAAGCGCACCCAGGACTTCTGGAGGTGCACTGGGCAATCCCCACACCCCGCGGCTGA 973  
Db 1988 ACAAGCGCACCCAGGACTTCTGGAGGTGCACTGGGCAATCCCCACACCCCGCGGCTGA 2047  
Qy 974 AGAAGAGAGAGCGTGACCGTGTGGAACGTGGGCGAGCGCTACTTCAGCGTGCCCTGG 1033  
Db 2048 AGAAGAGAGAGCGTGACCGTGTGGAACGTGGGCGAGCGCTACTTCAGCGTGCCCTGG 2107  
Qy 1034 ACAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAGACCCCG 1093  
Db 2108 ACAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAGACCCCG 2167  
Qy 1094 GCATCCGCTACAGTACAAACGTGTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153  
Db 2168 GCATCCGCTACAGTACAAACGTGTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2227  
Qy 1154 AGAGCAGATGACCAAGATCTTGAGAGCGCTTCCGCGCCCGCAACCCCGAGATCTGTATCT 1213  
Db 2228 AGAGCAGATGACCAAGATCTTGAGAGCGCTTCCGCGCCCGCAACCCCGAGATCTGTATCT 2287  
Qy 1214 ACCAGTACATGGAGCAGCTGTACGTGGGACGACCTGGAGATCGGCAGCACCGCGCCA 1273  
Db 2288 ACCA-----GGCCCCCTGTACGTGGGACGACCTGGAGATCGGCAGCACCGCGCCA 2341  
Qy 1274 AGATCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCCACACCCCGCAAGAAGC 1333  
Db 2342 AGATCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCCACACCCCGCAAGAAGC 2401  
Qy 1334 ACAGAGAGGAGCCCCCTTCTGTGTGATGGGCTAGGAGCTGCACCCCGCAGTAAGTGGACCG 1393  
Db 2402 ACCAGAGAGAGCCCCCTTCTGTGCCAT-----CGAGCTGACACCCCGCAAGTGGACCG 2455  
Qy 1394 TGCAGCCCATCGAGCTGCGGAGAGAGAGCTGGAACGTGAAGACATCCAGAAAGCTGG 1453  
Db 2456 TGCAGCCCATCGAGCTGCGGAGAGAGAGCTGGAACGTGAAGACATCCAGAAAGCTGG 2515  
Qy 1454 TGGGCAAGCTGAACCTGGGCCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCA 1513  
Db 2516 TGGGCAAGCTGAACCTGGGCCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCA 2575  
Qy 1514 AGCTGTGCGGCGCGCAAGGCCCTGACCGACATGTGTCGCCCTGACCGAGAGGCCGAGC 1573  
Db 2576 AGCTGTGCGGCGCGCAAGGCCCTGACCGACATGTGTCGCCCTGACCGAGAGGCCGAGC 2635  
Qy 1574 TGGAGCTGGCGGAGAACCGGAGATCTGCGGAGCGCGTGCACGGCGTGTACTACGACC 1633  
Db 2636 TGGAGCTGGCGGAGAACCGGAGATCTGCGGAGCGCGTGCACGGCGTGTACTACGACC 2695  
Qy 1634 CCAGCAAGGAGCTGGTGGCGAGATCCAGAAAGAGGCGCACGACAGTGGACCTTACCAGA 1693  
Db 2696 CCAGCAAGGAGCTGGTGGCGAGATCCAGAAAGAGGCGCACGACAGTGGACCTTACCAGA 2755  
Qy 1694 TCTACCAAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753  
Db 2756 TCTACCAAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 2815  
Qy 1754 ACACCAAGCAGTGAAGCAGTACCGAGCGGTGCAAGAGATCGCCATGGAGAGCATCG 1813  
Db 2816 ACACCAAGCAGTGAAGCAGTACCGAGCGGTGCAAGAGATCGCCATGGAGAGCATCG 2875

Qy 1814 TGATCTGGGCAAGACCCCCCAAGTTTCGCCCTGCCCATCTCAGAAAGAGACCTTGGGAGACCT 1873  
Db 2876 TGATCTGGGCAAGACCCCCCAAGTTTCGCCCTGCCCATCTCAGAAAGAGACCTTGGGAGACCT 2935  
Qy 1874 GGTGGACCGACTACTTGGCAGGCCACCTGGATCCCGGAGTGGAGTTCTGTGAACACCCCCC 1933  
Db 2936 GGTGGACCGACTACTTGGCAGGCCACCTGGATCCCGGAGTGGAGTTCTGTGAACACCCCCC 2995  
Qy 1934 CCCTGTGAAGCTGTGTATCCAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCT 1993  
Db 2996 CCCTGTGAAGCTGTGTATCCAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCT 3055  
Qy 1994 ACGTGGACGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 2053  
Db 3056 ACGTGGACGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACC 3115  
Qy 2054 GGGCCCGCAGAGATCGTGAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 2113  
Db 3116 GGGCCCGCAGAGATCGTGAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 3175  
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGAGGAGTGAACATCGTGACCGACAGCCAGT 2173  
Db 3176 CCATCCAGCTGGCCCTGCAGGACAGCGGAGGAGTGAACATCGTGACCGACAGCCAGT 3235  
Qy 2174 AGCCCTTGGGCAATCATCCAGGCCAGCCCGCAAGAGCGAGAGGCTGTGAAACCA 2233  
Db 3236 AGCCCTTGGGCAATCATCCAGGCCAGCCCGCAAGAGCGAGCTGTGAAACCA 3295  
Qy 2234 TCATCGAGCAGCTGATCAAGAAAGAGAGTGTACCTGAGCTGGGTGCCCGCCCAAGG 2293  
Db 3296 TCATCGAGCAGCTGATCAAGAAAGAGAGTGTACCTGAGCTGGGTGCCCGCCCAAGG 3355  
Qy 2294 GCATCGGCGCAACGAGCAGATCGACAAAGCTGGTGAAGCAAGGGCATCCGCAAGGTGCTGT 2353  
Db 3356 GCATCGGCGCAACGAGCAGATCGACAAAGCTGGTGAAGCAAGGGCATCCGCAAGGTGCTGT 3415  
Qy 2354 TCCTGGAACGGCATCGATGGCGGCATCGTGATCTTACCAAGTACATGGAACGACTGTACGTGG 2413  
Db 3416 TCCTGGAACGGCATCGATGGCGGCATCGTGATCTTACCAAGTACATGGAACGACTGTACGTGG 3475  
Qy 2414 GCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463  
Db 3476 GCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3525

Search completed: June 2, 2005, 04:23:32

Job time : 1424.77 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 10:29:57 ; Search time 1411.33 Seconds  
(without alignments)  
10728.225 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463

Sequence: 1 gtcagccaccatggccga.....gggctagcaccggtagaattc 2463

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5706582 seqs, 3073711274 residues

Total number of hits satisfying chosen parameters: 11413164

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2463	100.0	2463	10	US-09-899-575-31 Sequence 31, Appl
2	2455.4	99.7	2457	15	US-10-190-435-44 Sequence 44, Appl
3	2455.4	99.7	2457	16	US-10-190-305A-38 Sequence 38, Appl
4	2442.2	99.2	2469	10	US-09-899-575-30 Sequence 30, Appl
5	2436.2	98.9	2457	10	US-09-899-575-32 Sequence 32, Appl
6	2430.2	98.7	2457	15	US-10-190-435-45 Sequence 45, Appl
7	2430.2	98.7	2457	16	US-10-190-305A-39 Sequence 39, Appl
8	2422.6	98.4	2445	15	US-10-190-435-43 Sequence 43, Appl
9	2422.6	98.4	2445	16	US-10-190-305A-37 Sequence 37, Appl
10	2415.6	98.1	3930	15	US-10-190-435-9 Sequence 9, Appl
11	2414	98.0	3930	15	US-10-190-435-10 Sequence 10, Appl

12	2414	98.0	3930	15	US-10-190-435-11	Sequence 11, Appl
13	2414	98.0	5184	15	US-10-190-435-58	Sequence 58, Appl
14	2414	98.0	5184	16	US-10-190-305A-82	Sequence 82, Appl
15	2383.6	96.8	3531	15	US-10-190-435-13	Sequence 13, Appl
16	2382	96.7	3537	15	US-10-190-435-14	Sequence 14, Appl
17	2382	96.7	3537	15	US-10-190-435-15	Sequence 15, Appl
18	2381	96.7	5145	15	US-10-190-435-12	Sequence 12, Appl
19	2381	96.7	5145	16	US-10-190-305A-12	Sequence 12, Appl
20	2322.6	94.3	3607	15	US-10-190-435-48	Sequence 48, Appl
21	2322.6	94.3	3607	16	US-10-190-305A-42	Sequence 42, Appl
22	2322.6	94.3	3624	15	US-10-190-435-47	Sequence 47, Appl
23	2322.6	94.3	3624	16	US-10-190-305A-41	Sequence 41, Appl
24	2304.4	93.6	3597	15	US-10-190-435-46	Sequence 46, Appl
25	2304.4	93.6	3597	16	US-10-190-305A-40	Sequence 40, Appl
26	2142	87.0	2466	16	US-10-241-009-31	Sequence 31, Appl
27	2142	87.0	2466	16	US-10-190-434B-31	Sequence 31, Appl
28	2142	87.0	2466	16	US-10-190-305A-33	Sequence 33, Appl
29	2122.8	86.2	2472	16	US-10-241-009-32	Sequence 32, Appl
30	2122.8	86.2	2472	16	US-10-190-434B-32	Sequence 32, Appl
31	2122.8	86.2	2472	16	US-10-190-305A-34	Sequence 34, Appl
32	2115.2	85.9	2460	16	US-10-241-009-30	Sequence 30, Appl
33	2115.2	85.9	2460	16	US-10-190-434B-30	Sequence 30, Appl
34	2115.2	85.9	2460	16	US-10-190-305A-32	Sequence 32, Appl
35	2114.2	85.8	3564	16	US-10-241-009-13	Sequence 13, Appl
36	2114.2	85.8	3564	16	US-10-241-009-14	Sequence 14, Appl
37	2114.2	85.8	3564	16	US-10-190-434B-13	Sequence 13, Appl
38	2114.2	85.8	3564	16	US-10-190-434B-14	Sequence 14, Appl
39	2113.6	85.8	4716	15	US-10-190-435-17	Sequence 17, Appl
40	2113.6	85.8	4716	16	US-10-190-305A-13	Sequence 13, Appl
41	2109.8	85.7	3999	16	US-10-241-009-9	Sequence 9, Appl
42	2109.8	85.7	3999	16	US-10-190-434B-9	Sequence 9, Appl
43	2108.2	85.6	3999	16	US-10-241-009-10	Sequence 10, Appl
44	2108.2	85.6	3999	16	US-10-241-009-11	Sequence 11, Appl
45	2108.2	85.6	3999	16	US-10-190-434B-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-899-575-31  
; Sequence 31, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megede, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Estrelita Janse  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; CURRENT FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 09/475,704  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 31  
; LENGTH: 2463  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PR975YM  
US-09-899-575-31

Query Match 100.0%; Score 2463; DB 10; Length 2463;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GTCAGCCACCACCTGGCCGAGGCCATGAGCCAGCCAGCCAGCCATCTGTGAG 60  
Db 1 GTCAGCCACCACCTGGCCGAGGCCATGAGCCAGCCAGCCAGCCATCTGTGAG 60



QY 61 CGCAGCAATTCAAGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGC 120  
 DB |||||  
 QY 61 CGCAGCAATTCAAGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGC 120  
 DB |||||  
 QY 121 CATATGCCCCGAACCTGCGCGCCCCCGCGCAAGAGGCTGTGGAGTGCAGCAAGGAG 180  
 DB |||||  
 QY 121 CATATGCCCCGAACCTGCGCGCCCCCGCGCAAGAGGCTGTGGAGTGCAGCAAGGAG 180  
 DB |||||  
 QY 181 GGCACACAGATGAAGGACTGCAACGAGCGCCAGGCCCAATCTTCTTCGCGAGGACCTGGCC 240  
 DB |||||  
 QY 181 GGCACACAGATGAAGGACTGCAACGAGCGCCAGGCCCAATCTTCTTCGCGAGGACCTGGCC 240  
 DB |||||  
 QY 241 TTCCCCCAGGCAAGGCCCCCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACACGCCACC 300  
 DB |||||  
 QY 241 TTCCCCCAGGCAAGGCCCCCGAGTTCCTCCAGCGAGGAGAACCGCGCAACACGCCACC 300  
 DB |||||  
 QY 301 AGCCGCGAGTGCAGGTGCGCGCGCAAAACCCCGCAGCGAGGCGCGCGCGAGCCAG 360  
 DB |||||  
 QY 301 AGCCGCGAGTGCAGGTGCGCGCGCAAAACCCCGCAGCGAGGCGCGCGCGAGCCAG 360  
 DB |||||  
 QY 361 GGCACCTGAACTTCCCCCAGATCACCTGTGTGCGAGGCCCCCTGTGTAGCATCAAGGTG 420  
 DB |||||  
 QY 361 GGCACCTGAACTTCCCCCAGATCACCTGTGTGCGAGGCCCCCTGTGTAGCATCAAGGTG 420  
 DB |||||  
 QY 421 GCGCGCAGATCAAGGAGGCCCTGTGTGACACCGCGCGCAGACACACCGTGTGGAGGAG 480  
 DB |||||  
 QY 421 GCGCGCAGATCAAGGAGGCCCTGTGTGACACCGCGCGCAGACACACCGTGTGGAGGAG 480  
 DB |||||  
 QY 481 ATGAGGCTGCCCCGGAAGTGAAGCCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAG 540  
 DB |||||  
 QY 481 ATGAGGCTGCCCCGGAAGTGAAGCCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAG 540  
 DB |||||  
 QY 541 GTGCGCGAGTACACACAGATCTCTGATCGAGATCTGCGGCGCAAGAGGCGCATCGCGCAG 600  
 DB |||||  
 QY 541 GTGCGCGAGTACACACAGATCTCTGATCGAGATCTGCGGCGCAAGAGGCGCATCGCGCAG 600  
 DB |||||  
 QY 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 DB |||||  
 QY 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 DB |||||  
 QY 661 ACCCTGAACTTCCCAATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG 720  
 DB |||||  
 QY 661 ACCCTGAACTTCCCAATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG 720  
 DB |||||  
 QY 721 GACGCCCCCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 780  
 DB |||||  
 QY 721 GACGCCCCCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 780  
 DB |||||  
 QY 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCCCCGAGAACCCCTAC 840  
 DB |||||  
 QY 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCCCCGAGAACCCCTAC 840  
 DB |||||  
 QY 841 AACACCCCGTGTTCGCCATCAAGAGAGAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA 900  
 DB |||||  
 QY 841 AACACCCCGTGTTCGCCATCAAGAGAGAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA 900  
 DB |||||  
 QY 901 TTCGCGAGCTGAACAGCGCAACCCAGGACTTCTGGAGGTGTCAGCTGGGCGATCCCCAC 960  
 DB |||||  
 QY 901 TTCGCGAGCTGAACAGCGCAACCCAGGACTTCTGGAGGTGTCAGCTGGGCGATCCCCAC 960  
 DB |||||  
 QY 961 CCCGCGGCTGAAGAGAGAGAGCGTGAACCGTGTGGAAGTGGCGAGCGCTACTTTC 1020  
 DB |||||  
 QY 961 CCCGCGGCTGAAGAGAGAGAGCGTGAACCGTGTGGAAGTGGCGAGCGCTACTTTC 1020  
 DB |||||  
 QY 1021 AGCGTGCCCTTGACGAGGACTTCCGCAAGTACACCGCCCTTCAACATCCCCAGCATCAAC 1080  
 DB |||||  
 QY 1021 AGCGTGCCCTTGACGAGGACTTCCGCAAGTACACCGCCCTTCAACATCCCCAGCATCAAC 1080  
 DB |||||  
 QY 1081 AACGAGACCCCGGCTACCGTACAGTACAAAGTGTGCTGCCCCAGGGCTGGAGGGGAGC 1140  
 DB |||||  
 QY 1081 AACGAGACCCCGGCTACCGTACAGTACAAAGTGTGCTGCCCCAGGGCTGGAGGGGAGC 1140  
 DB |||||  
 QY 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCC 1200  
 DB |||||

DB 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC 1200  
 QY |||||  
 QY 1201 GAGATCGTATCTACAGGCCCCCTGTAGCTGGGCGAGCGACTGGAGATCGGCGAGCAC 1260  
 DB |||||  
 QY 1201 GAGATCGTATCTACAGGCCCCCTGTAGCTGGGCGAGCGACTGGAGATCGGCGAGCAC 1260  
 DB |||||  
 QY 1261 CGCGCCAAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTCAACACCCCGAC 1320  
 DB |||||  
 QY 1261 CGCGCCAAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTCAACACCCCGAC 1320  
 DB |||||  
 QY 1321 AAGAAGCACCCAGAGAGGCCCTTCTTGTGTGATGGGCTTACGAGCTGCAACCCCGACAAG 1380  
 DB |||||  
 QY 1321 AAGAAGCACCCAGAGAGGCCCTTCTTGTGTGATGGGCTTACGAGCTGCAACCCCGACAAG 1380  
 DB |||||  
 QY 1381 TGGACCGTGCAGCCATCGAGCTGCCGAGAGAGAGAGCTGGAACCTGAACGATCCAG 1440  
 DB |||||  
 QY 1381 TGGACCGTGCAGCCATCGAGCTGCCGAGAGAGAGAGCTGGAACCTGAACGATCCAG 1440  
 DB |||||  
 QY 1441 AAGCTGCTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCGCGCATCAAGGTGCGCGAG 1500  
 DB |||||  
 QY 1441 AAGCTGCTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCGCGCATCAAGGTGCGCGAG 1500  
 DB |||||  
 QY 1501 CTGTGCAAGCTGCTGCGCGCGCCAAAGGCCCTGAACGACATCGTGCCCTGACCCGAGGAG 1560  
 DB |||||  
 QY 1501 CTGTGCAAGCTGCTGCGCGCGCCAAAGGCCCTGAACGACATCGTGCCCTGACCCGAGGAG 1560  
 DB |||||  
 QY 1561 GCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGGCCCTGCAACGCGCTGTAC 1620  
 DB |||||  
 QY 1561 GCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGGCCCTGCAACGCGCTGTAC 1620  
 DB |||||  
 QY 1621 TAGCACCCAGCAAGGACCTGTGTGGCCGAGATCTCAGAAGCAGGCGCCACGACAGTGGACC 1680  
 DB |||||  
 QY 1621 TAGCACCCAGCAAGGACCTGTGTGGCCGAGATCTCAGAAGCAGGCGCCACGACAGTGGACC 1680  
 DB |||||  
 QY 1681 TACCAGATCTACGAGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCGAAGATGCGC 1740  
 DB |||||  
 QY 1681 TACCAGATCTACGAGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCGAAGATGCGC 1740  
 DB |||||  
 QY 1741 ACGGCCACACCAACGATGAAGCAGCTGACCGAGGCGTGCAGAGATCGGCATCGGAG 1800  
 DB |||||  
 QY 1741 ACGGCCACACCAACGATGAAGCAGCTGACCGAGGCGTGCAGAGATCGGCATCGGAG 1800  
 DB |||||  
 QY 1801 AGCATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACTGG 1860  
 DB |||||  
 QY 1801 AGCATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACTGG 1860  
 DB |||||  
 QY 1861 GAGACTGTGTGAACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1920  
 DB |||||  
 QY 1861 GAGACTGTGTGAACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1920  
 DB |||||  
 QY 1921 ACGCCCCCTGTGTGAAGCTGTGGTACCGAGCTGGAGAGGCCATCATCCGCGCGGAG 1980  
 DB |||||  
 QY 1921 ACGCCCCCTGTGTGAAGCTGTGGTACCGAGCTGGAGAGGCCATCATCCGCGCGGAG 1980  
 DB |||||  
 QY 1981 ACCTTCTAGTGAAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGCTG 2040  
 DB |||||  
 QY 1981 ACCTTCTAGTGAAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGCTG 2040  
 DB |||||  
 QY 2041 ACCGACCGGGCGCGGAGAGTCTGTAGCTTGAACCGCAGACCAACCAAGAGAGAGAG 2100  
 DB |||||  
 QY 2041 ACCGACCGGGCGCGGAGAGTCTGTAGCTTGAACCGCAGACCAACCAAGAGAGAGAG 2100  
 DB |||||  
 QY 2101 CTGCGAGCCATCCAGCTGGCCCTGCGAGGAGGCGGCGAGGAGTGAACATCTGTGACCGAC 2160  
 DB |||||  
 QY 2101 CTGCGAGCCATCCAGCTGGCCCTGCGAGGAGGCGGCGAGGAGTGAACATCTGTGACCGAC 2160  
 DB |||||  
 QY 2161 AGCCAGTACGCTGCGGCTCATCCAGGCGCCAGCCGCAAGAGCGAGGAGCGAGCTGGTG 2220  
 DB |||||  
 QY 2161 AGCCAGTACGCTGCGGCTCATCCAGGCGCCAGCCGCAAGAGCGAGGAGCGAGCTGGTG 2220  
 DB |||||  
 QY 2221 AACAGATCATCGAGCAGCTGATCAAGAGAGGAGAGAGTGTACTGTAGCTGGGTGCGCGCC 2280  
 DB |||||

```
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTTGAGTGGTGCCTGCC 2280
Qy 2281 CACAAGGCGCATCGCGGCAACAGCAGATGACAAAGCTGTGTAGCAAGGGCATCCGCAAG 2340
Db 2281 CACAAGGCGCATCGCGGCAACAGCAGATGACAAAGCTGTGTAGCAAGGGCATCCGCAAG 2340
Qy 2341 GTGCTGTCTTGGAGCGGCATCGATGCGGGCATCGTGATCTACCAAGTACATGAGCAGCTG 2400
Db 2341 GTGCTGTCTTGGAGCGGCATCGATGCGGGCATCGTGATCTACCAAGTACATGAGCAGCTG 2400
Qy 2401 TACGTGGCAGCGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCGGTGAA 2460
Db 2401 TACGTGGCAGCGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCGGTGAA 2460
Qy 2461 TTC 2463
Db 2461 TTC 2463

RESULT 2
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGERDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44

Query Match 99.7%; Score 2455.4; DB 15; Length 2457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCAGCCGACCATGCGCGAGGCCATGAGCCAGGCCACAGCGCCAAACATCCTGATGAG 60
Db 1 GTCAGCCGACCATGCGCGAGGCCATGAGCCAGGCCACAGCGCCAAACATCCTGATGAG 60
Qy 61 CGCAGCAACTTCAGGGGCCCCAGGGCATCATCAAGTGTCTCACTGGGCGAGGGGC 120
Db 61 CGCAGCAACTTCAGGGGCCCCAGGGCATCATCAAGTGTCTCACTGGGCGAGGGGC 120
Qy 121 CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGGCAAGGAG 180
Db 121 CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGGCAAGGAG 180
Qy 181 GGCACACAGATGAAGGACTGACCGAGCGCGGCCGCAAGCTTCTTCGCGAGGACCTGGCC 240
Db 181 GGCACACAGATGAAGGACTGACCGAGCGCGGCCGCAAGCTTCTTCGCGAGGACCTGGCC 240
Qy 241 TTCCCCCAGGCGAGGCCCGGGAGTTCCCGAGCGAGCAGAACCGGGCCACAGCCCCCACC 300
Db 241 TTCCCCCAGGCGAGGCCCGGGAGTTCCCGAGCGAGCAGAACCGGGCCACAGCCCCCACC 300
Qy 301 AGCCGCGAGCTGCAGGTGCGGGGCAACACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360
Db 301 AGCCGCGAGCTGCAGGTGCGGGGCAACACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360
Qy 361 GGCACCCCTGAACCTTCCCCCAGATCACCCCTGTGGCAGCGGCCCTTGGTGAGCATCAAGGTG 420
```

```
Db 361 GGCACCCCTGAACCTTCCCCCAGATCACCCCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTG 420
Qy 421 GGCGGCCAGATCAAGGAGGCCCTCTGTGACACCGCGCGCCGACGACACCGTGTCTGGAGGAG 480
Db 421 GGCGGCCAGATCAAGGAGGCCCTCTGTGACACCGCGCGCCGACGACACCGTGTCTGGAGGAG 480
Qy 481 ATGAGCCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG 540
Db 481 ATGAGCCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG 540
Qy 541 GTGCGCAGATACGACGATCCTGATCGAGATCTCGCGCAAGAGGCCCATCGGCACCGGTG 600
Db 541 GTGCGCAGTACGACGATCCTGATCGAGATCTCGCGCAAGAGGCCCATCGGCACCGGTG 600
Qy 601 CTGATCGGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGGTGC 660
Db 601 CTGATCGGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGGTGC 660
Qy 661 ACCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGCATG 720
Db 661 ACCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGCATG 720
Qy 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCCGAGGAGAGATCAAGGCCCTTGACCGCC 780
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCCGAGGAGAGATCAAGGCCCTTGACCGCC 780
Qy 781 ATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGCGCCCCCGAGAACCCCTTAC 840
Db 781 ATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGCGCCCCCGAGAACCCCTTAC 840
Qy 841 AACACCCCCGTGTTTCGCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCCGTGTTTCGCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGGCATCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGGCATCCCCAC 960
Qy 961 CCGCGCGGCTTGAAGAAGAAGAGCGTGACCGTGTGACGCTGGGCGAGCGCTACTTTC 1020
Db 961 CCGCGCGGCTTGAAGAAGAAGAGCGTGACCGTGTGACGCTGGGCGAGCGCTACTTTC 1020
Qy 1021 AGCGTGCCCTTGGACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCGAGCATCAAC 1080
Db 1021 AGCGTGCCCTTGGACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCGAGCATCAAC 1080
Qy 1081 AACGAGACCCCGGCATCCGCTACCACTACCACTGCTGCCCGCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCATCCGCTACCACTACCACTGCTGCCCGCAGGGCTGGAAGGGCAGC 1140
Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCCGCGCGCCGCAACCCCC 1200
Db 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCCGCGCGCCGCAACCCCC 1200
Qy 1201 GAGATCGTGATCTACCAAGGCCCTCTGTAGTGGGCGAGCGACCTGGAGATCGGGCAGCAC 1260
Db 1201 GAGATCGTGATCTACCAAGGCCCTCTGTAGTGGGCGAGCGACCTGGAGATCGGGCAGCAC 1260
Qy 1261 CGCGCAGAGATCGAGGAGCTGCGCAGCACCTGTGCTGGCTGGGGCTTCAACACCCCGGAC 1320
Db 1261 CGCGCAGAGATCGAGGAGCTGCGCAGCACCTGTGCTGGCTGGGGCTTCAACACCCCGGAC 1320
Qy 1321 AAGAAGCACAGAGGAGGCCCTCTCTGTGATGGGTACGAGCTGCACCCCGACAAAG 1380
Db 1321 AAGAAGCACAGAGGAGGCCCTCTCTGTGATGGGTACGAGCTGCACCCCGACAAAG 1380
Qy 1381 TGGACCGTGCAGGCCCTCAGCTGCCCCGAGAGGAGAGCTGGACCGGTGAACGACATCCAG 1440
Db 1381 TGGACCGTGCAGGCCCTCAGCTGCCCCGAGAGGAGAGCTGGACCGGTGAACGACATCCAG 1440
Qy 1441 AAGTGTGGGCAAGCTGAACCTGGGCGAGCGAGATCTACCCCGGCATCAAGGTGGCGAG 1500
Db 1441 AAGTGTGGGCAAGCTGAACCTGGGCGAGCGAGATCTACCCCGGCATCAAGGTGGCGAG 1500
```

```
Db 1441 AAGCTGTGGCAAGCTGAACCTGGCCAGCCAGATCTACCCCGGCATCAAGGTGGCCAG 1500
Qy 1501 CTGTGCAAGTGTCTGGCGCGCCCAAGCCCTGACCGAATCTGTGCCCTTGACCGAGGAG 1560
Db 1501 CTGTGCAAGTGTCTGGCGCGCCCAAGCCCTGACCGAATCTGTGCCCTTGACCGAGGAG 1560
Qy 1561 GCCGAGTGTAGTGTGGCCGAGAACCCGAGATCTCTGGCGAGCCCGTGACGCGCTGTAC 1620
Db 1561 GCCGAGTGTAGTGTGGCCGAGAACCCGAGATCTCTGGCGAGCCCGTGACGCGCTGTAC 1620
Qy 1621 TACGACCCCAAGCAAGCACTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAGTGAC 1680
Db 1621 TACGACCCCAAGCAAGCACTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAGTGAC 1680
Qy 1681 TACGAGATCTACGAGGCGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGTGCGC 1740
Db 1681 TACGAGATCTACGAGGCGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGTGCGC 1740
Qy 1741 ACCGCCCAACCAAGCAAGCTGAAGCAGCTGACCGAGCCGTGCAGAAGATCGCATGGAG 1800
Db 1741 ACCGCCCAACCAAGCAAGCTGAAGCAGCTGACCGAGCCGTGCAGAAGATCGCATGGAG 1800
Qy 1801 AGCATCGTGTCTGTGGGCGAAGACCCCAAGTTCGGCTTGCCCATCCAGAAGAGACCTGG 1860
Db 1801 AGCATCGTGTCTGTGGGCGAAGACCCCAAGTTCGGCTTGCCCATCCAGAAGAGACCTGG 1860
Qy 1861 GAGACCTGTGTGACCGACTACTGTGGAGCCACCTGTGATCCCGCAGTGGGATTCGTGAAC 1920
Db 1861 GAGACCTGTGTGACCGACTACTGTGGAGCCACCTGTGATCCCGCAGTGGGATTCGTGAAC 1920
Qy 1921 ACCCCCCCTGTGAAGCTGTGTGACCGAGTGTGACCGAGGAGCCCATCATCGCGCCGAG 1980
Db 1921 ACCCCCCCTGTGTGAAGCTGTGTGACCGAGTGTGACCGAGGAGCCCATCATCGCGCCGAG 1980
Qy 1981 ACCTTCTACGTGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTAGCTG 2040
Db 1981 ACCTTCTACGTGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTAGCTG 2040
Qy 2041 ACCGACCGGGGCGCGCAAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Db 2041 ACCGACCGGGGCGCGCAAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Qy 2101 CTGACGCCATCCAGCTGGCCCTGACGAGCAGCGGAGGAGTGAACATCTGTGACCGAC 2160
Db 2101 CTGACGCCATCCAGCTGGCCCTGACGAGCAGCGGAGGAGTGAACATCTGTGACCGAC 2160
Qy 2161 AGCCAGTACGCCCTGGGCGATCATCCAGGCCAGCCGACAAAGAGCGAGCGAGCTGGTG 2220
Db 2161 AGCCAGTACGCCCTGGGCGATCATCCAGGCCAGCCGACAAAGAGCGAGCGAGCTGGTG 2220
Qy 2221 AACCATGATCATGACGAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCC 2280
Db 2221 AACCATGATCATGACGAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCC 2280
Qy 2281 CACAAGGGCATCGCGGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGGCATCCGCAAG 2340
Db 2281 CACAAGGGCATCGCGGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGGCATCCGCAAG 2340
Qy 2341 GTGCTGTCTCGACCGGCACTGATGGCGGATCGTGTATCTACAGTACATGACGACCTG 2400
Db 2341 GTGCTGTCTCGACCGGCACTGATGGCGGATCGTGTATCTACAGTACATGACGACCTG 2400
Qy 2401 TAGTGGGCGAGCGCGCCCTAGATGATTAAGCTTCCCGGGGTAGCACCGGT 2457
Db 2401 TAGTGGGCGAGCGCGCCCTAGATGATTAAGCTTCCCGGGGTAGCACCGGT 2457
```

RESULT 3  
US-10-190-305A-38  
; Sequence 38, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MESEDE, Jan

```
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; FILE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190,305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C  
US-10-190-305A-38
```

```
Query Match 99.7%; Score 2455.4; DB 16; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GTCGAGCCACCATGGCCGAGGCGCATGAGCGAGGCCACAGCGCCCAACATCTCTGATGCAG 60  
Db 1 GTCGAGCCACCATGGCCGAGGCGCATGAGCGAGGCCACAGCGCCCAACATCTCTGATGCAG 60  
Qy 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGGC 120  
Db 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGGC 120  
Qy 121 CACATCGCGCAACTGCGCGCGCCCGCAAGAGGGCTGTGTGAAGTGCGGCAAGGAG 180  
Db 121 CACATCGCGCAACTGCGCGCGCCCGCAAGAGGGCTGTGTGAAGTGCGGCAAGGAG 180  
Qy 181 GGCACACAGATGAAGACTGTGACCGAGCGCGAGGCCAACTTCTTCGCGAGGACCTGGCC 240  
Db 181 GGCACACAGATGAAGACTGTGACCGAGCGCGAGGCCAACTTCTTCGCGAGGACCTGGCC 240  
Qy 241 TTCTCCCGAGGCGAAGCGCGCGAGTTCCTCCAGCGAGCAACCGCGCAACAGCCCGCAC 300  
Db 241 TTCTCCCGAGGCGAAGCGCGCGAGTTCCTCCAGCGAGCAACCGCGCAACAGCCCGCAC 300  
Qy 301 AGCGCGAGTGTGAGGTGGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGAG 360  
Db 301 AGCGCGAGTGTGAGGTGGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGAG 360  
Qy 361 GGCACCTTGAATTTCTCCCGCAGATCACCTGTGTGCGAGCGCGCGCGCGCGCGCGAG 420  
Db 361 GGCACCTTGAATTTCTCCCGCAGATCACCTGTGTGCGAGCGCGCGCGCGCGCGCGAG 420  
Qy 421 GCGCGCGAGATCAAGGAGCGCTGTGTGGACACCGCGCGCGCGCGCGCGCGCGCGAG 480  
Db 421 GCGCGCGAGATCAAGGAGCGCTGTGTGGACACCGCGCGCGCGCGCGCGCGCGCGAG 480  
Qy 481 ATGAGCCTTCCCGCAAGTGAAGCCAAAGATGATCGCGGCGCATCGCGCGCTTTCATCAAG 540  
Db 481 ATGAGCCTTCCCGCAAGTGAAGCCAAAGATGATCGCGGCGCATCGCGCGCTTTCATCAAG 540  
Qy 541 GTGCGCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACGCTG 600  
Db 541 GTGCGCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACGCTG 600  
Qy 601 CTGATCGGCGCCACCGCGTGAACATCATCTGCGCGCAACATCTGCTGACCGAGTGGCTGC 660  
Db 601 CTGATCGGCGCCACCGCGTGAACATCATCTGCGCGCAACATCTGCTGACCGAGTGGCTGC 660  
Qy 661 ACCCTGAACCTTCCCGCATCGAGACCGTTCGCGCGCGCGCGCGCGCGCGCGAG 720  
Db 661 ACCCTGAACCTTCCCGCATCGAGACCGTTCGCGCGCGCGCGCGCGCGCGCGAG 720  
Qy 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTTGACCGCC 780  
Db 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTTGACCGCC 780
```



QY	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGGC	120
Db			
QY	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGGC	120
Db			
QY	121	CACATCGCCGCAACTCGCCGCCCCCGCCGCAAGAAGGGCTGTGGAAGTGC CGCAAGGAG	180
Db			
QY	121	CACATCGCCGCAACTCGCCGCCCCCGCCGCAAGAAGGGCTGTGGAAGTGC CGCAAGGAG	180
Db			
QY	181	GGCCACAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGAGACTGTGGCC	240
Db			
QY	181	GGCCACAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGAGACTGTGGCC	240
Db			
QY	241	TTCCGCCAGGGCAAGGCCCGCGAGTTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Db			
QY	241	TTCCGCCAGGGCAAGGCCCGCGAGTTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Db			
QY	301	AGCCGGAGTGTGAGGTGCGCGCGCAACACCCCGCAGCGAGGCGCGCGCGAGGCCAG	360
Db			
QY	301	AGCCGGAGTGTGAGGTGCGCGCGCAACACCCCGCAGCGAGGCGCGCGCGAGGCCAG	360
Db			
QY	361	GGCACCCTGAACTTCCGCCAGATACCGCTGTGCGAGGGCCCCCTGTGTAGCATCAAGGTG	420
Db			
QY	361	GGCACCCTGAACTTCCGCCAGATACCGCTGTGCGAGCGCCCCCTGTGTAGCATCAAGGTG	420
Db			
QY	421	GGCGGCAGATCAAGGAGGCCCTGTCTGGACACCGCGCGCGACGACACCGTGTGGAAGGAG	480
Db			
QY	421	GGCGGCAGATCAAGGAGGCCCTGTCTGGACACCGCGCGCGACGACACCGTGTGGAAGGAG	480
Db			
QY	481	ATGAGCTGTCGCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG	540
Db			
QY	481	ATGAGCTGTCGCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG	540
Db			
QY	541	GTGCGCGAGTACGACAGATCCTGTGATCGAGATCTGCGGCGAGAGGCGCATCGGCAACGTG	600
Db			
QY	541	GTGCGCGAGTACGACAGATCCTGTGATCGAGATCTGCGGCGAGAGGCGCATCGGCAACGTG	600
Db			
QY	601	CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGTGACCCAGCTGGGCTGC	660
Db			
QY	601	CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGTGACCCAGCTGGGCTGC	660
Db			
QY	661	ACCCTGAACTTCCCATCAAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG	720
Db			
QY	661	ACCCTGAACTTCCCATCAAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG	720
Db			
QY	721	GACGCCCCAAGTGAAGTGAAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCCCTGACCGCC	780
Db			
QY	721	GACGCCCCAAGTGAAGTGAAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCCCTGACCGCC	780
Db			
QY	781	ATCTGCGAGGAGTGAAGGAGGAGGCGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTAC	840
Db			
QY	781	ATCTGCGAGGAGTGAAGGAGGAGGCGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTAC	840
Db			
QY	841	AACAACCCCGGTGTTCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Db			
QY	841	AACAACCCCGGTGTTCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Db			
QY	901	TTCCGGGAGCTGAACAGCGCACCCAGGACTTCTGAGGAGTGCAGCTGGGCATCCCCCAC	960
Db			
QY	901	TTCCGGGAGCTGAACAGCGCACCCAGGACTTCTGAGGAGTGCAGCTGGGCATCCCCCAC	960
Db			
QY	961	CCCGCCGGCTGAAGAAAGAAAGAGCGTGACCGTGTGGAAGTGGCGACGCGCTACTTC	1020
Db			
QY	961	CCCGCCGGCTGAAGAAAGAAAGAGCGTGACCGTGTGGAAGTGGCGACGCGCTACTTC	1020
Db			
QY	1021	AGCGTCCCTGGACGAGGACTTCCGCAAGTACACCGCCCTTCAACATCCCCAGCATCAAC	1080
Db			
QY	1021	AGCGTCCCTGGACGAGGACTTCCGCAAGTACACCGCCCTTCAACATCCCCAGCATCAAC	1080
Db			
QY	1081	AACGAGACCCCGGCTACCGCTACCAAGTACAAAGTGTGCTGCTGCCCCAGGGCTGGAGGGGAGC	1140
Db			
QY	1081	AACGAGACCCCGGCTACCGCTACCAAGTACAAAGTGTGCTGCTGCCCCAGGGCTGGAGGGGAGC	1140
Db			

QY	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCGCGCCGCGCAACCCC	1200
Db			
QY	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCGCGCCGCGCAACCCC	1200
Db			
QY	1201	GAGATCGTGATCTACCA-----GGCCCCCTGTAGCTGGGCGAGCGACTCTGGAGATCGGC	1254
Db			
QY	1201	GAGATCGTGATCTACCAAGTACATGAGCGACCTGTAGCTGGGCGAGCGACTCTGGAGATCGGC	1260
Db			
QY	1255	GAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGCTTTCACACC	1314
Db			
QY	1261	GAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGCTTTCACACC	1320
Db			
QY	1315	CCCGACAAGAAGCACCAGAAGGAGGCCCTTCTCTGTGATGGGCTACGAGCTGCAACCCC	1374
Db			
QY	1321	CCCGACAAGAAGCACCAGAAGGAGGCCCTTCTCTGTGATGGGCTACGAGCTGCAACCCC	1380
Db			
QY	1375	GACAACTGAGCCGTGAGCCCATCGAGCTGCGCGAGAGGAGAGCTGAGACCTGTGAACGAC	1434
Db			
QY	1381	GACAACTGAGCCGTGAGCCCATCGAGCTGCGCGAGAGGAGCTGAGACCTGTGAACGAC	1440
Db			
QY	1435	ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCC CGGCATCAAGGTG	1494
Db			
QY	1441	ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCC CGGCATCAAGGTG	1500
Db			
QY	1495	CGCCAGCTGTGAAGCTGTGCGCGCGCAAGGCCCTTGAACGACATCTGTGCCCTTGACC	1554
Db			
QY	1501	CGCCAGCTGTGAAGCTGTGCGCGCGCAAGGCCCTTGAACGACATCTGTGCCCTTGACC	1560
Db			
QY	1555	GAGGAGCCGAGCTGAGCTGCGCGAGAACCGGAGATCTGTGGCGAGCCCTGTGACGGC	1614
Db			
QY	1561	GAGGAGCCGAGCTGAGCTGCGCGAGAACCGGAGATCTGTGGCGAGCCCTGTGACGGC	1620
Db			
QY	1615	GTGTACTACGACCCCGAGCAAGGACCTGTGTGGCGCGAGATCCAGAAGCAGGGCCACGACCAG	1674
Db			
QY	1621	GTGTACTACGACCCCGAGCAAGGACCTGTGTGGCGCGAGATCCAGAAGCAGGGCCACGACCAG	1680
Db			
QY	1675	TGGACCTACAGATCTACGAGGAGCCCTTCAAGAACTCTGAAGACCCGCGCAAGTACGCCAAG	1734
Db			
QY	1681	TGGACCTACAGATCTACGAGGAGCCCTTCAAGAACTCTGAAGACCCGCGCAAGTACGCCAAG	1740
Db			
QY	1735	ATGCGCACCGCCCAACCAACGAGTGAAGCAGCTGACCGAGGCCCTGTGCAAGATCGCC	1794
Db			
QY	1741	ATGCGCACCGCCCAACCAACGAGTGAAGCAGCTGACCGAGGCCCTGTGCAAGATCGCC	1800
Db			
QY	1795	ATGAGAGGATCTGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAG	1854
Db			
QY	1801	ATGAGAGGATCTGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAG	1860
Db			
QY	1855	ACCTGGGAGACCTGTGTGACCGACTACTGTGGCAGGCCACTTGGATCCCCAGTGGGAGTTC	1914
Db			
QY	1861	ACCTGGGAGACCTGTGTGACCGACTACTGTGGCAGGCCACTTGGATCCCCAGTGGGAGTTC	1920
Db			
QY	1915	GTGAACACCCCGCCCTGTGTGAGCTGTGGTACCAGCTGGAGAGGAGGCCCATCATCGGC	1974
Db			
QY	1921	GTGAACACCCCGCCCGCTGTGTGAGCTGTGGTACCAGCTGGAGAGGAGGCCCATCATCGGC	1980
Db			
QY	1975	GCGGAGACCTTCTACCTGTGAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGCGC	2034
Db			
QY	1981	GCGGAGACCTTCTACCTGTGAGCGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGCGC	2040
Db			
QY	2035	TAGCTGACCGAGCCGGGGCGCGCAAGAGATCTGTGAGCTTGAACCGAGACCAACCAAGAG	2094
Db			
QY	2041	TAGCTGACCGAGCCGGGGCGCGCAAGAGATCTGTGAGCTTGAACCGAGACCAACCAAGAG	2100
Db			
QY	2095	ACGAGCTGAGGCCCATTCAGCTGGCCCTGCGAGGACGCGGCGAGCGAGTGAACATCGTG	2154
Db			
QY	2101	ACGAGCTGAGGCCCATTCAGCTGGCCCTGCGAGGACGCGGCGAGCGAGTGAACATCGTG	2160
Db			
QY	2155	ACCGAGCAGCAGTACGCGCCCTGGGCATCATCCAGGCCCGAGCCCGACAAGAGCGAGAGCGAG	2214
Db			
QY	2161	ACCGAGCAGCAGTACGCGCCCTGGGCATCATCCAGGCCCGAGCCCGACAAGAGCGAGAGCGAG	2220
Db			
QY	2215	CTGGTGAAACCGAGATCATCGAGCAGCTGTATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG	2274
Db			

Db 2221 CTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2280  
Qy |||||  
Db 2275 CCOCGCCACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334  
Qy |||||  
Db 2281 CCOCGCCACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340  
Qy |||||  
Db 2335 GGCNAGGTGCTGTTCTCTGGAGCGGATCGATCGCGGCAATCGTGATCTACCAAGTACATGAC 2394  
Qy |||||  
Db 2341 GGCNAGGTGCTGTTCTCTGGAGCGGATCGATCGCGGCAATCGTGATCTACCAAGTACATGAC 2400  
Qy |||||  
Db 2395 GACCTGTACGTGGGACGCGGCGGCTTAGGATCGATTAAAGCTTTCCCGGGCTAGCAC 2454  
Qy |||||  
Db 2401 GACCTGTACGTGGGACGCGGCGGCTTAGGATCGATTAAAGCTTTCCCGGGCTAGCAC 2460  
Qy |||||  
Db 2455 GGTGAATTC 2463  
Qy |||||  
Db 2461 GGTGAATTC 2469

## RESULT 5

US-09-899-575-32

; Sequence 32, Application US/09899575

; Publication No. US20030223961A1

; GENERAL INFORMATION:

; APPLICANT: Zur Megede, Jan

; APPLICANT: Barnett, Susan W.

; APPLICANT: Egnebrecht, Susan

; APPLICANT: van Rensburg, Estrelita Jause

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C

; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

; FILE REFERENCE: PP01631.102

; CURRENT APPLICATION NUMBER: US/09/899,575

; CURRENT FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: 09/475,704

; PRIOR FILING DATE: 1999-12-30

; NUMBER OF SEQ ID NOS: 135

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 32

; LENGTH: 2457

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: PR975YMM

US-09-899-575-32

Query Match 98.9%; Score 2436.2; DB 10; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy 1 GTCGACGCCACCATGGCGGCGGCGCATGAGCCAGGCCACAGCGCCAAACATCTGATGCAG 60  
Db |||||  
Qy 1 GTCGACGCCACCATGGCGGCGGCGCATGAGCCAGGCCACAGCGCCAAACATCTGATGCAG 60  
Db |||||  
Qy 61 GCGAGCAATTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120  
Db |||||  
Qy 61 GCGAGCAATTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120  
Db |||||  
Qy 121 CACATCGCCGCAACTGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db |||||  
Qy 121 CACATCGCCGCAACTGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db |||||  
Qy 181 GGCACACAGATGAAGGACTGACCGAGCGCGCAGGCCAACTTTCTTCGCGAGGAGCCTGGCC 240  
Db |||||  
Qy 181 GGCACACAGATGAAGGACTGACCGAGCGCGCAGGCCAACTTTCTTCGCGAGGAGCCTGGCC 240  
Db |||||  
Qy 241 TTCCCCCAGGGCAAGGCCCGGAGTTCCCCAGCGAGCAGAACCGCGCCAAACAGCCCCCACC 300  
Db |||||  
Qy 241 TTCCCCCAGGGCAAGGCCCGGAGTTCCCCAGCGAGCAGAACCGCGCCAAACAGCCCCCACC 300  
Db |||||  
Qy 301 AGCGCGAGCTGCAGGTGCGGGCGCAACACCCCGCAGCGGCCCGCGCGCGAGCCAG 360  
Db |||||  
Qy 301 AGCGCGAGCTGCAGGTGCGGGCGCAACACCCCGCAGCGGCCCGCGCGCGAGCCAG 360  
Db |||||



1441 AAGCTGTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGCATCAAGGTGGCCAG 1500  
 1435 AAGCTGTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGCATCAAGGTGGCCAG 1494  
 1501 CTGTGCAAGCTGTGGGCGGCGCAAGGCGCTGACCGACATCTGTGCCCTTGACCGAGGAG 1560  
 1495 CTGTGCAAGCTGTGGGCGGCGCAAGGCGCTGACCGACATCTGTGCCCTTGACCGAGGAG 1554  
 1561 GCCGAGCTGGAGTGGGCGGAGAACCGGAGATCTGTGGGAGCCCTGTGCAAGGCTGTAC 1620  
 1555 GCCGAGCTGGAGTGGGCGGAGAACCGGAGATCTGTGGGAGCCCTGTGCAAGGCTGTAC 1614  
 1621 TAGACACCCAGCAAGCACTGTGTGGGCGGAGATCCAGAGCAGGCGCACACCAAGTGGACC 1680  
 1615 TAGACACCCAGCAAGCACTGTGTGGGCGGAGATCCAGAGCAGGCGCACACCAAGTGGACC 1674  
 1681 TACAGATCTTACAGAGGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCG 1740  
 1675 TACAGATCTTACAGAGGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCG 1734  
 1741 ACCGCCACACCAACGAGTGAAGCAGCTGACCGAGGCGCTGCAGAGATCGCCATGGAG 1800  
 1735 ACCGCCACACCAACGAGTGAAGCAGCTGACCGAGGCGCTGCAGAGATCGCCATGGAG 1794  
 1801 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCGGCTGCCCATCCAGAGGAGACTGTG 1860  
 1795 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCGGCTGCCCATCCAGAGGAGACTGTG 1854  
 1861 GAGACTGTGTGACCGACTACTGTGGCGAGCCACTGTGATCCCGAGTGGAGTTCGTGAAC 1920  
 1855 GAGACTGTGTGACCGACTACTGTGGCGAGCCACTGTGATCCCGAGTGGAGTTCGTGAAC 1914  
 1921 ACCCCCCCTGTGTGAAGTGTGTGACCGAGTGGAGGAGCCCATCTCCGCGCGAG 1980  
 1915 ACCCCCCCTGTGTGAAGTGTGTGACCGAGTGGAGGAGCCCATCTCCGCGCGAG 1974  
 1981 ACCTTCTAGCTGACGCGCGCGCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTG 2040  
 1975 ACCTTCTAGCTGACGCGCGCGCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTG 2034  
 2041 ACCGACCGGGGCGGAGAGATCTGTGAGCCTGACCGGAGACCAACCAAGAGACCGAG 2100  
 2035 ACCGACCGGGGCGGAGAGATCTGTGAGCCTGACCGGAGACCAACCAAGAGACCGAG 2094  
 2101 CTGAGGCGCATTCAGCTGGCCCTGAGGACAGGCGGAGGAGTGAACATCGTGACCGAC 2160  
 2095 CTGAGGCGCATTCAGCTGGCCCTGAGGACAGGCGGAGGAGTGAACATCGTGACCGAC 2154  
 2161 AGCCAGTACGCTGGGCGATCATCCAGGCGCGGAGCAAGAGCGAGCGAGCTGGTG 2220  
 2155 AGCCAGTACGCTGGGCGATCATCCAGGCGCGGAGCAAGAGCGAGCGAGCTGGTG 2214  
 2221 AACAGATCATTCAGAGCTGTATCAAGAGGAGAGGTGTA CTTGAGCTGGGTGCCCGCC 2280  
 2215 AACAGATCATTCAGAGCTGTATCAAGAGGAGAGGTGTA CTTGAGCTGGGTGCCCGCC 2274  
 2281 CACAGGGGATTCGGCGGCAAGCAGATCGACAACTGTGTGAGCAAGGGCATCCGCAAG 2340  
 2275 CACAGGGGATTCGGCGGCAAGCAGATCGACAACTGTGTGAGCAAGGGCATCCGCAAG 2334  
 2341 GTGCTGTCTCTGACCGGCAATCGATGGCGGATCGTATCTACCACTATGACGACCTTG 2400  
 2335 GTGCTGTCTCTGACCGGCAATCGATGGCGGATCGTATCTACCACTATGACGACCTTG 2394  
 2401 TACGTGGGACCGCGGCGCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
 2395 TACGTGGGACCGCGGCGCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2454  
 2461 TTC 2463  
 2455 TTC 2457

RESULT 6  
 US-10-190-435-45  
 ; Sequence 45, Application US/10190435  
 ; Publication No. US20030143248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEDEDE, Jan  
 ; APPLICANT: BARNETT, Susan W.  
 ; APPLICANT: LIAN, Ying  
 ; APPLICANT: ENGELBRECHT, Susan  
 ; APPLICANT: VAN RENSBURG, Estrelita J.  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
 ; FILE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
 ; FILE REFERENCE: PP18133.003 / 2302-18133  
 ; CURRENT APPLICATION NUMBER: US/10/190,435  
 ; CURRENT FILING DATE: 2002-12-30  
 ; NUMBER OF SEQ ID NOS: 319  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 45  
 ; LENGTH: 2457  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt\_C  
 US-10-190-435-45

Query Match 98.7%; Score 2430.2; DB 15; Length 2457;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
 QY 7 GCCACATCGGCGAGGCCATGAGCCAGGCCACGAGGCCAACATCTGTGATGCGAGCGCAGC 66  
 DB 1 GCCACATCGGCGAGGCCATGAGCCAGGCCACGAGGCCAACATCTGTGATGCGAGCGCAGC 60  
 QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCGAGGAGGCCACATC 126  
 DB 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCGAGGAGGCCACATC 120  
 QY 127 GCCCGCAACTGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTGCAGGAGGAGGGCCAC 186  
 DB 121 GCCCGCAACTGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTGCAGGAGGAGGGCCAC 180  
 QY 187 CAGATGAAGGACTGCAACCGAGGCCAGGCCAACTTTCTTCGCGAGGACCTGGGCTTCCCC 246  
 DB 181 CAGATGAAGGACTGCAACCGAGGCCAGGCCAACTTTCTTCGCGAGGACCTGGGCTTCCCC 240  
 QY 247 CAGGGCAAGGCGCGCAGTTCCTCCAGCGAGCGAGCAACCGGCCAACAGCCCCACAGCCGC 306  
 DB 241 CAGGGCAAGGCGCGCAGTTCCTCCAGCGAGCGAGCAACCGGCCAACAGCCCCACAGCCGC 300  
 QY 307 GAGCTCAGGTGCGCGGCGCAACACCCCGCAGCGAGGCGCGGCGCCAGCGCCAGGGCACC 366  
 DB 301 GAGCTCAGGTGCGCGGCGCAACACCCCGCAGCGAGGCGCGGCGCCAGCGCCAGGGCACC 360  
 QY 367 CTGAACCTTCCCCCAGATCAACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGCGGC 426  
 DB 361 CTGAACCTTCCCCCAGATCAACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGCGGC 420  
 QY 427 CAGATCAAGAGGCGCCCTGTGTGACACCGGGCGCGACGACACCGTGTGGAGGAGATGAGC 486  
 DB 421 CAGATCAAGAGGCGCCCTGTGTGACACCGGGCGCGACGACACCGTGTGGAGGAGATGAGC 480  
 QY 487 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAGGTGGCG 546  
 DB 481 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAGGTGGCG 540  
 QY 547 CAGTACGACACAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGACACCGTGTGATC 606  
 DB 541 CAGTACGACACAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGACACCGTGTGATC 600  
 QY 607 GGGCCCAACCCCGTGAACATCATCGGCGCGCAATCTGACCCAGCTGGGCTGCACCCCTG 666  
 DB 601 GGGCCCAACCCCGTGAACATCATCGGCGCGCAATCTGACCCAGCTGGGCTGCACCCCTG 660

```
Qy 667 AACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATCGACGGC 726
Db 661 AACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATCGACGGC 720
Qy 727 CCCAAGGTGAAGAGTGGCCCTCAGACCGAGAGAGATCAAGAGCCCTGACCGCCATCTGC 786
Db 721 CCCAAGGTGAAGAGTGGCCCTCAGACCGAGAGAGATCAAGAGCCCTGACCGCCATCTGC 780
Qy 787 GAGGAGATGGAGAGAGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTCAACAACC 846
Db 781 GAGGAGATGGAGAGAGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTCAACAACC 840
Qy 847 CCCGTGTTCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
Db 841 CCCGTGTTCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy 907 GAGCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966
Db 901 GAGCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Qy 967 GGCCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
Db 961 GGCCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Qy 1027 CCCCTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
Db 1021 CCCCTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy 1087 ACCCCCGGCATCCCGTACCAAGTCAAAAGTCTGCGCCCGAGAGAGAGAGAGAGAGAGAG 1146
Db 1081 ACCCCCGGCATCCCGTACCAAGTCAAAAGTCTGCGCCCGAGAGAGAGAGAGAGAGAGAG 1140
Qy 1147 ATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
Db 1141 ATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Qy 1207 GTGATCTACCA-----GGCCCCCTGTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 GTGATCTACCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1261 CGGSCCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 CGGSCCAAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 1321 AAGAAGCACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 AAGAAGCACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy 1441 AAGCTGTGGGCAAGCTGAACTGGGGCCAGCAGATCTACCCCGCATCAAGAGTGGCGCAG 1500
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGGCCAGCAGATCTACCCCGCATCAAGAGTGGCGCAG 1500
Qy 1501 CTGTGCAAGTGTCTGCGCGGCGCAAGCCCTGACCGACATCTGCGCCCTGACCGAGAGAG 1560
Db 1501 CTGTGCAAGTGTCTGCGCGGCGCAAGCCCTGACCGACATCTGCGCCCTGACCGAGAGAG 1560
Qy 1561 GCCGAGCTGAGCTGGCGCGAGAACCCGAGATCTCTGCGCGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 1561 GCCGAGCTGAGCTGGCGCGAGAACCCGAGATCTCTGCGCGAGAGAGAGAGAGAGAGAGAGAG 1620
Qy 1621 TACGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Db 1621 TACGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Qy 1681 TACCAGATCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db 1681 TACCAGATCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy 1741 ACCGCCCACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
```

```
Db 1741 ACCGCCCACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy 1801 AGCATCTGTGATCTGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 1801 AGCATCTGTGATCTGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy 1861 GAGACCTGTGTGAGACCGGACTACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 1861 GAGACCTGTGTGAGACCGGACTACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy 1921 ACCCCCGCCCTGGTGAAGCTGTGTGATCGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 ACCCCCGCCCTGGTGAAGCTGTGTGATCGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy 1981 ACCTTCTACGTGGAAGCGCGCCCAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db 1981 ACCTTCTACGTGGAAGCGCGCCCAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Qy 2041 ACCGACCGGGCGCGGCAAGAGATCGTGAGCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2041 ACCGACCGGGCGCGGCAAGAGATCGTGAGCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Qy 2101 CTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2101 CTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Qy 2161 AGCCAGTACCGCCCTGGGCGCATCATCCAGCGCCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db 2161 AGCCAGTACCGCCCTGGGCGCATCATCCAGCGCCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy 2221 AACAGATATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2221 AACAGATATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Qy 2281 CACAAGGGCATCGGCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db 2281 CACAAGGGCATCGGCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy 2341 GTGCTGTCTTGGACCGGCATCGATGCGGCATCGTATCTACAGTACATCGAGAGAGAGAGAGAG 2400
Db 2341 GTGCTGTCTTGGACCGGCATCGATGCGGCATCGTATCTACAGTACATCGAGAGAGAGAGAGAG 2400
Qy 2401 TACGTGGGCAAGCGCGCCCTTAGGATCGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2457
Db 2401 TACGTGGGCAAGCGCGCCCTTAGGATCGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2457
```

## RESULT 7

```
US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MESEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39
```

```
Query Match          98.7%; Score 2430.2; DB 16; Length 2457;
Best Local Similarity 99.6%; Pred. No. 0;
```

Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;	
Qy	7 GCCACCATGGCCGAGGCGCATGAGCCAGGCCACAGCGCCCAACATCTGTGATGCGAGCCGAGC 66
Db	
Qy	1 GCCACCATGGCCGAGGCGCATGAGCCAGGCCACAGCGCCCAACATCTGTGATGCGAGCCGAGC 60
Db	
Qy	67 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db	
Qy	61 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATC 120
Db	
Qy	127 GCCCGCAACTGCGCGCCCGCCCGGCAAGAGGGCTGTGTGAAGTGTGCGGCAAGAGGGCCAC 186
Db	
Qy	121 GCCCGCAACTGCGCGCCCGCCCGGCAAGAGGGCTGTGTGAAGTGTGCGGCAAGAGGGCCAC 180
Db	
Qy	187 CAGATGAAGGACTGCAACGAGCGCCAGGCCAACTTCTTCGCGAGAGACCTGGCTTCCCC 246
Db	
Qy	181 CAGATGAAGGACTGCAACGAGCGCCAGGCCAACTTCTTCGCGAGAGACCTGGCTTCCCC 240
Db	
Qy	247 CAGGGCAAGCCCGCGAGTTCCCGCAGCGAGCAGAAACCGCGCCCAACAGCCCCACACGCGC 306
Db	
Qy	241 CAGGGCAAGCCCGCGAGTTCCCGCAGCGAGCAGAAACCGCGCCCAACAGCCCCACACGCGC 300
Db	
Qy	307 GAGCTGAGGTGCGCGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 366
Db	
Qy	301 GAGCTGAGGTGCGCGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 360
Db	
Qy	367 CTGAACTTCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGCGGC 426
Db	
Qy	361 CTGAACTTCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGCGGC 420
Db	
Qy	427 CAGATCAAGGAGGCGCTGTGTGACACCGGCGCGCAACACCGTGTGTGAGGAGATGAGC 486
Db	
Qy	421 CAGATCAAGGAGGCGCTGTGTGACACCGGCGCGCAACACCGTGTGTGAGGAGATGAGC 480
Db	
Qy	487 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGCTTTCATCAAGGTGCGC 546
Db	
Qy	481 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGCTTTCATCAAGGTGCGC 540
Db	
Qy	547 CAGTAGACCATGATCTGATCGAGATCTGGGCAAGAGGCCATCGGCAACCGTGTGATC 606
Db	
Qy	541 CAGTAGACCATGATCTGATCGAGATCTGGGCAAGAGGCCATCGGCAACCGTGTGATC 600
Db	
Qy	607 GGGCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666
Db	
Qy	601 GGGCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660
Db	
Qy	667 AACTTCCCATCAGCCCATCGAGACCGTGCCTGTGAAGTGAAGCGCCGCGCATGAGCGGC 726
Db	
Qy	661 AACTTCCCATCAGCCCATCGAGACCGTGCCTGTGAAGTGAAGCGCCGCGCATGAGCGGC 720
Db	
Qy	727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGTACCGGCCATCTGC 786
Db	
Qy	721 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGTACCGGCCATCTGC 780
Db	
Qy	787 GAGGAGATGAGAGGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTCAACACC 846
Db	
Qy	781 GAGGAGATGAGAGGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTCAACACC 840
Db	
Qy	847 CCCGTGTTCCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGAGCTTCCGC 906
Db	
Qy	841 CCCGTGTTCCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGAGCTTCCGC 900
Db	
Qy	907 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGGCATCCCCCAACCCGCGC 966
Db	
Qy	901 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGGCATCCCCCAACCCGCGC 960
Db	
Qy	967 GGCCTGAAGAAAGAGAGCGGTGACCGTCTGGAAGTGGGCGACCGCTACTTTCAGCGTG 1026
Db	
Qy	961 GGCCTGAAGAAAGAGAGCGGTGACCGTCTGGAAGTGGGCGACCGCTACTTTCAGCGTG 1020
Db	
Qy	1027 CCCCTGGAACGAGACTTCCGCAAGTACACCGCCTTACCAATCCCCAGCATCAACAGAG 1086
Db	
Qy	1021 CCCCTGGAACGAGACTTCCGCAAGTACACCGCCTTACCAATCCCCAGCATCAACAGAG 1080
Db	

Qy	1087 ACCCCGCGCATCGCTACCACTACAGTGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCCAGC 1146
Db	
Qy	1081 ACCCCGCGCATCGCTACCACTACAGTGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCCAGC 1140
Db	
Qy	1147 ATCTTCCAGAGCAGCATGATCAAGATCTCTGGAGCCCTTTCGCGCCCGCAACCCCGAGATC 1206
Db	
Qy	1141 ATCTTCCAGAGCAGCATGATCAAGATCTCTGGAGCCCTTTCGCGCCCGCAACCCCGAGATC 1200
Db	
Qy	1207 GTGATCTACCA-----GGCCCCCTGTAGTGGGCGAGCGACCTGGAGATCGGCCAGCAC 1260
Db	
Qy	1201 GTGATCTACCACTACATGATGACGACCTGTAGTGGGCGAGCGACCTGGAGATCGGCCAGCAC 1260
Db	
Qy	1261 CGCGCCAAAGATCGAGGAGCTTGGCAAGCACCTCTCTGCTGTGGGGCTTTCACACCCCGGAC 1320
Db	
Qy	1261 CGCGCCAAAGATCGAGGAGCTTGGCAAGCACCTCTCTGCTGTGGGGCTTTCACACCCCGGAC 1320
Db	
Qy	1321 AAGAAACAACAGAAAGAGCCCTTCTCTGTGTGGATGGGTCTACAGCTGCACCCCGACAAG 1380
Db	
Qy	1321 AAGAAACAACAGAAAGAGCCCTTCTCTGTGTGGATGGGTCTACAGCTGCACCCCGACAAG 1380
Db	
Qy	1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACCGTGCAGCATCCAG 1440
Db	
Qy	1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACCGTGCAGCATCCAG 1440
Db	
Qy	1441 AAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db	
Qy	1441 AAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db	
Qy	1501 CTGTGCAAGCTGTGTGCGCGCGCAAGGGCCCTGACCGAATCTGTGCCCTTGAACCGAGGAG 1560
Db	
Qy	1501 CTGTGCAAGCTGTGTGCGCGCGCAAGGGCCCTGACCGAATCTGTGCCCTTGAACCGAGGAG 1560
Db	
Qy	1561 GCGGAGCTGGAGCTGGCGGAGACCGGAGATCTCTGGGAGCGCCCTGTGCACCGCGTGTAC 1620
Db	
Qy	1561 GCGGAGCTGGAGCTGGCGGAGACCGGAGATCTCTGGGAGCGCCCTGTGCACCGCGTGTAC 1620
Db	
Qy	1621 TAGCACCCAGCAAGGACCTGTGTGGCGCGAGATCTCAGAAAGCAGGGGCCACCAAGTGTGACC 1680
Db	
Qy	1621 TAGCACCCAGCAAGGACCTGTGTGGCGCGAGATCTCAGAAAGCAGGGGCCACCAAGTGTGACC 1680
Db	
Qy	1681 TACAGATCTTACCGAGGCGCTTCAAGAACCTTGAAGACCGGCGCAAGTACGCAAGATGCGC 1740
Db	
Qy	1741 ACCGCGCACCAACCAACCGTGAAGCAGCTGACCGAGGCGGTGAGAGATCGCATGAG 1800
Db	
Qy	1741 ACCGCGCACCAACCAACCGTGAAGCAGCTGACCGAGGCGGTGAGAGATCGCATGAG 1800
Db	
Qy	1801 AGCATCTGTATCTGGGCGCAAGACCCCCCAAGTTCCGCTTGCCTTCCCATCCAGAAAGAGACCTGG 1860
Db	
Qy	1801 AGCATCTGTATCTGGGCGCAAGACCCCCCAAGTTCCGCTTGCCTTCCCATCCAGAAAGAGACCTGG 1860
Db	
Qy	1861 GAGACTGTGTGAGACCACTACTGGCAGGCGCACCTGGATCTCCAGTGGGAGTTCGTGAAC 1920
Db	
Qy	1861 GAGACTGTGTGAGACCACTACTGGCAGGCGCACCTGGATCTCCAGTGGGAGTTCGTGAAC 1920
Db	
Qy	1921 ACCCCCCCTGTGTGAAGCTGTGTACCACTGAGGAGGAGGCCATCATCGGCGCGGAG 1980
Db	
Qy	1921 ACCCCCCCTGTGTGAAGCTGTGTACCACTGAGGAGGAGGCCATCATCGGCGCGGAG 1980
Db	
Qy	1981 ACCTTCTTCTGTGAGCGGCGCGCAACCCCGCAGAACCAAGATCGGCAAGGCGCGGTACGTG 2040
Db	
Qy	1981 ACCTTCTTCTGTGAGCGGCGCGCAACCCCGCAGAACCAAGATCGGCAAGGCGCGGTACGTG 2040
Db	
Qy	2041 ACCGACCGGGGCGCGCAGAAAGTCTGTAGCTTACCGAGACCAACCAAGAGAGAGAGAG 2100
Db	
Qy	2041 ACCGACCGGGGCGCGCAGAAAGTCTGTAGCTTACCGAGACCAACCAAGAGAGAGAGAG 2100
Db	
Qy	2101 CTGCAAGGCGCATCTGAGTGGCCCTGCAAGGACAGCGGCGAGGAGTGAACATCTGTGACCGAC 2160
Db	
Qy	2101 CTGCAAGGCGCATCTGAGTGGCCCTGCAAGGACAGCGGCGAGGAGTGAACATCTGTGACCGAC 2160
Db	

```
Qy 2161 AGCCAGTACGCTGGGTCATCATCAGGCGCCAGCCCGACAAGAGCGAGCGAGCTGGTG 2220
Db 2161 AGCCAGTACGCTGGGTCATCATCAGGCGCCAGCCCGACAAGAGCGAGCGAGCTGGTG 2220
Qy 2221 AACAGATCATCGAGCAGCTGATCAAGAAGAGAGAGTGTACCTCAGCTGGGTGCCGCC 2280
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAAGAGAGAGTGTACCTCAGCTGGGTGCCGCC 2280
Qy 2281 CACAAGGCGATCGGCGGCAAGAGAGATCGAAGCTGGTGAGCAAGGGGATCGGCAAG 2340
Db 2281 CACAAGGCGATCGGCGGCAAGAGAGATCGAAGCTGGTGAGCAAGGGGATCGGCAAG 2340
Qy 2341 GTGCTGTTCTCGGAGCGGATCGATGGCGGCATCGTGATCTACCAAGTACATGAGACCTG 2400
Db 2341 GTGCTGTTCTCGGAGCGGATCGATGGCGGCATCGTGATCTACCAAGTACATGAGACCTG 2400
Qy 2401 TACGTGGCAGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457
Db 2401 TACGTGGCAGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 8
US-10-190-435-43
; Sequence 43: Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Esrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 43
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMMW_C
US-10-190-435-43

Query Match 98.4%; Score 2422.6; DB 15; Length 2445;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAAACATCCTTGATCGAGCGCAGC 66
Db 1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGCGCCAAACATCCTTGATCGAGCGCAGC 60
Qy 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGGCAAGGGGCCACATC 126
Db 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGGCAAGGGGCCACATC 120
Qy 127 GCCCGCAACTCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGGGGCCAC 186
Db 121 GCCCGCAACTCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGGGGCCAC 180
Qy 187 CAGATGAAGGACTGCACCGAGCGCCAGGCCCAACTTTCTTCGCGAGGAGCTTGGCTTCCCC 246
Db 181 CAGATGAAGGACTGCACCGAGCGCCAGGCCCAACTTTCTTCGCGAGGAGCTTGGCTTCCCC 240
Qy 247 CAGGGCAAGGCCCGCGAGTTCCCGAGGAGCAGAAACCGCGCAACAGCCCCCAGCGCGC 306
Db 241 CAGGGCAAGGCCCGCGAGTTCCCGAGGAGCAGAAACCGCGCAACAGCCCCCAGCGCGC 300
Qy 307 GAGCTGCAGGTGCGGGCGCAACCCCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 366
Db 301 GAGCTGCAGGTGCGGGCGCAACCCCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 360
```

```
Qy 367 CTGAATTTCCCCCAGATCAACCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTGGCGGC 426
Db 361 CTGAATTTCCCCCAGATCAACCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTGGCGGC 420
Qy 427 CAGATCAAGGAGGCCCTCTGTGGAACAACCGCGCCGACGACACCGTCTCGAGGAGATGAGC 486
Db 421 CAGATCAAGGAGGCCCTCTGTGGAACAACCGCGCCGACGACACCGTCTCGAGGAGATGAGC 480
Qy 487 CTGCCCCGCGAAGTGAAGCCCAAGATGATCGCGCGGCATCTCGCGGCTTTCATCAAGGTGGC 546
Db 481 CTGCCCCGCGAAGTGAAGCCCAAGATGATCGCGCGGCATCTCGCGGCTTTCATCAAGGTGGC 540
Qy 547 CAGTACGACAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATC 606
Db 541 CAGTACGACAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATC 600
Qy 607 GGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666
Db 601 GGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660
Qy 667 AACTTCCCATTCAGCCCCCATCGAGACCGTGTGCCGTGAAGCTGAAGCCCGGATGAGCGGC 726
Db 661 AACTTCCCATTCAGCCCCCATCGAGACCGTGTGCCGTGAAGCTGAAGCCCGGATGAGCGGC 720
Qy 727 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGC 786
Db 721 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGC 780
Qy 787 GAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGNACCCCTACNACACC 846
Db 781 GAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGNACCCCTACNACACC 840
Qy 847 CCCGTGTTTGGCCATCAAGAAGAAGAGCAGCACCAGTGGCGCCAAAGCTGGTGACTTCCGC 906
Db 841 CCCGTGTTTGGCCATCAAGAAGAAGAGCAGCACCAGTGGCGCCAAAGCTGGTGACTTCCGC 900
Qy 907 GAGCTGAACAAGCGCAACCCAGGACTTCTGGGAGGTGAGTGTGGGATCCCCCAACCCCGCC 966
Db 901 GAGCTGAACAAGCGCAACCCAGGACTTCTGGGAGGTGAGTGTGGGATCCCCCAACCCCGCC 960
Qy 967 GGCCTGAAGAAGAGAGAGGCTGACCGTGTGAGCTGGGCGAGCGCTACTTCAGCGTG 1026
Db 961 GGCCTGAAGAAGAGAGAGGCTGACCGTGTGAGCTGGGCGAGCGCTACTTCAGCGTG 1020
Qy 1027 CCCCTGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCGCAGCATCAACAACGAG 1086
Db 1021 CCCCTGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCGCAGCATCAACAACGAG 1080
Qy 1087 ACCCCCGGCATTCGCTACCAAGTACAACTGTGCTGCCCAAGGGCTGGAAGGGCAGCCCCAGC 1146
Db 1081 ACCCCCGGCATTCGCTACCAAGTACAACTGTGCTGCCCAAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCTGTGAGCGCTTTCGCGCGCCGCAACCCCGAGTC 1206
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCTGTGAGCGCTTTCGCGCGCCGCAACCCCGAGTC 1200
Qy 1207 GTGATCTTACCAAGGCCCGCTGTGTGAGGAGCAGCAGCTGAGAGATCGGCGCAGCAGCGCGCC 1266
Db 1201 GTGATCTTACCAAGGCCCGCTGTGTGAGGAGCAGCAGCTGAGAGATCGGCGCAGCAGCGCGCC 1260
Qy 1267 AAGATCGAGGAGCTGCGCAAGCAGCCTGCTGCGCTGGGGCTTTCACCAACCCCGCAAGAAG 1326
Db 1261 AAGATCGAGGAGCTGCGCAAGCAGCCTGCTGCGCTGGGGCTTTCACCAACCCCGCAAGAAG 1320
Qy 1327 CACCAAGAGGAGCGCCCTTCTGTGTGATGGGCTACGAGCTGCGACCCCGCAAGTGGACC 1386
Db 1321 CACCAAGAGGAGCGCCCTTCTGTGTGATGGGCTTCTGTGATGGGCTTTCACCAACCCCGCAAGTGGACC 1374
Qy 1387 GTGCAGGCCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAAAGCTG 1446
Db 1375 GTGCAGGCCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAAAGCTG 1434
```

```
QY 1447 GTGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1506
DB |||||||
DB 1435 GTGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGC 1494
QY 1507 AAGCTGTGCGCGGCCAAGGCCCTTGACCGAATCTGTGCCCCCTGACCGAGGAGGCCGAG 1566
DB |||||||
DB 1495 AAGCTGTGCGCGGCCAAGGCCCTTGACCGAATCTGTGCCCCCTGACCGAGGAGGCCGAG 1554
QY 1567 CTGGAGTGGCCGAGAACCCGAGATCCTGCGGAGCCCGTGCACGGCGTGTAAGGAC 1626
DB |||||||
DB 1555 CTGGAGTGGCCGAGAACCCGAGATCCTGCGGAGCCCGTGCACGGCGTGTAAGGAC 1614
QY 1627 CCAGCAAGAACCTGTGTGGCCGAGATCCAGAAGCAGGGCCACACAGTGGACCTACCAG 1686
DB |||||||
DB 1615 CCAGCAAGAACCTGTGTGGCCGAGATCCAGAAGCAGGGCCACACAGTGGACCTACCAG 1674
QY 1687 ATCTACCGAGGACCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746
DB |||||||
DB 1675 ATCTACCGAGGACCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734
QY 1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCGAGAGATCGCCATGGAGAGATC 1806
DB |||||||
DB 1735 CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCGAGAGATCGCCATGGAGAGATC 1794
QY 1807 GTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAAGGAGACCTGGGAGACC 1866
DB |||||||
DB 1795 GTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAAGGAGACCTGGGAGACC 1854
QY 1867 TGTGGAACGACTACTGTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCTGTGAACACCCCC 1926
DB |||||||
DB 1855 TGTGGAACGACTACTGTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTCTGTGAACACCCCC 1914
QY 1927 CCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGCGCCGAGACCTTC 1986
DB |||||||
DB 1915 CCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGCGCCGAGACCTTC 1974
QY 1987 TAGCTGGAACGCGCCGCAACCCGAGACCAAGATCGGCAAGGCCCGCTACGTGACCGAC 2046
DB |||||||
DB 1975 TAGCTGGAACGCGCCGCAACCCGAGACCAAGATCGGCAAGGCCCGCTACGTGACCGAC 2034
QY 2047 CGGGGCCGCGAGAAAGATCGTGAAGCTTGAACCGAGACCAACCAAGAACCGAGCTGCAG 2106
DB |||||||
DB 2035 CGGGGCCGCGAGAAAGATCGTGAAGCTTGAACCGAGACCAACCAAGAACCGAGCTGCAG 2094
QY 2107 GCCATCCAGCTGGCCCTGCGAGGACAGCGGCGAGCGAGTGAACATCGTGACCGACGCCAG 2166
DB |||||||
DB 2095 GCCATCCAGCTGGCCCTGCGAGGACAGCGGCGAGCGAGTGAACATCGTGACCGACGCCAG 2154
QY 2167 TAGCCCTGGGCATCATCCAGGCCCAAGCCGACCAAGAGCGAGAGCGAGCTGGTGAACCCAG 2226
DB |||||||
DB 2155 TAGCCCTGGGCATCATCCAGGCCCAAGCCGACCAAGAGCGAGAGCGAGCTGGTGAACCCAG 2214
QY 2227 ATCATCCAGCAGCTGATCAAGAAAGGAGAGGTGTAACCTGAGCTGGGTGCCCCGCCCAAG 2286
DB |||||||
DB 2215 ATCATCCAGCAGCTGATCAAGAAAGGAGAGGTGTAACCTGAGCTGGGTGCCCCGCCCAAG 2274
QY 2287 GGCATCGGGGCCAACGAGCAGATCGAAGCTGGTGAAGCAGGGCATCGCAGAGTGTCTG 2346
DB |||||||
DB 2275 GGCATCGGGGCCAACGAGCAGATCGAAGCTGGTGAAGCAGGGCATCGCAGAGTGTCTG 2334
QY 2347 TTCTCTGAGCGGCATCGATGGCGGCATCGTGATCTACACAGTACATGACGACCTGTACGTG 2406
DB |||||||
DB 2335 TTCTCTGAGCGGCATCGATGGCGGCATCGTGATCTACAGTACATGAGCAGCTGTACGTG 2394
QY 2407 GGCAGCGGGGCCCTTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCCGT 2457
DB |||||||
DB 2395 GGCAGCGGGGCCCTTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCCGT 2445
```

RESULT 9

US-10-190-305A-37

; Sequence 37, Application US/10190305A

; Publication No. US20030198621A1

```
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: p2Pol.opt.YMMW_C
US-10-190-305A-37
```

Query Match 98.4%; Score 2422.6; DB 16; Length 2445;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
QY 7 GCCACCATGGCGAGGCCCATGAGCCAGGCCACCCAGCGCCAAACATCCTGATGCAAGCGCAGC 66
DB |||||||
DB 1 GCCACCATGGCGAGGCCCATGAGCCAGGCCACCCAGCGCCAAACATCCTGATGCAAGCGCAGC 60
QY 67 AACTTCAAGGGGCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGGCCACATC 126
DB |||||||
DB 61 AACTTCAAGGGGCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGGCCACATC 120
QY 127 GCCCGCAACTGCGCGCCCCCGCCAGCAAGAGGGCTGCTGGAAGTGCAGGAGGGGCCAC 186
DB |||||||
DB 121 GCCCGCAACTGCGCGCCCCCGCCAGCAAGAGGGCTGCTGGAAGTGCAGGAGGGGCCAC 180
QY 187 CAGATGAAGGACTGCACCGAGGCCACCGAGGCCAACTTCTTCGCGAGGACCTTGGCTTCCCC 246
DB |||||||
DB 181 CAGATGAAGGACTGCACCGAGGCCACCGAGGCCAACTTCTTCGCGAGGACCTTGGCTTCCCC 240
QY 247 CAGGGCAAGGCCCGCGAGTTCCCGAGCAGAGAAACCGGCCCAACAGCCCCACAGCGCGC 306
DB |||||||
DB 241 CAGGGCAAGGCCCGCGAGTTCCCGAGCAGAGAAACCGGCCCAACAGCCCCACAGCGCGC 300
QY 307 GAGCTGCAGGTGCGGGCGCACAAACCCCGCAGGAGGCCGCCCGCCAGCGCCAGGGCCACC 366
DB |||||||
DB 301 GAGCTGCAGGTGCGGGCGCACAAACCCCGCAGGAGGCCGCCCGCCAGCGCCAGGGCCACC 360
QY 367 CTGAACCTTCCCCCAGATCACCTCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTGGCGGC 426
DB |||||||
DB 361 CTGAACCTTCCCCCAGATCACCTCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTGGCGGC 420
QY 427 CAGATCAAGGAGGCCCTGTGTGACACCGGGCGCGACGACACCGTGTGTGAGGAGATGAGC 486
DB |||||||
DB 421 CAGATCAAGGAGGCCCTGTGTGACACCGGGCGCGACGACACCGTGTGTGAGGAGATGAGC 480
QY 487 CTGCCCCGCAAGTGGGAAGGCCCAAGATGATCGGGGCATCGGGGCTTCATCAAGGTGCGC 546
DB |||||||
DB 481 CTGCCCCGCAAGTGGGAAGGCCCAAGATGATCGGGGCATCGGGGCTTCATCAAGGTGCGC 540
QY 547 CAGTACGACAGATCTCTGATCCGAGATCTCGGCAAGAGGCCATCGGCACCGTGTGATC 606
DB |||||||
DB 541 CAGTACGACAGATCTCTGATCCGAGATCTCGGCAAGAGGCCATCGGCACCGTGTGATC 600
QY 607 GGGCCCCACCCCGTGAACATCATCTGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666
DB |||||||
DB 601 GGGCCCCACCCCGTGAACATCATCTGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660
QY 667 AACTTCCCCCATCAGCCCCCATCGNAGCCGTCGCTGAAGTGAAGTGAAGTGAAGTGAAGTGA 726
DB |||||||
DB 661 AACTTCCCCCATCAGCCCCCATCGNAGCCGTCGCTGAAGTGAAGTGAAGTGAAGTGAAGTGA 720
QY 727 CCCAAGGTGAAGCAGTGGGCCCTCTGACCGAGGAGAAAGATCAAGGCCCTTGACCGCCATCTGC 786
```

Db 721 CCCAAGGTGAAGCAGTGGCCCCCTGACGAGAGAGATCAAGGGCCCTGACCGCCATCTGC 780  
Qy 787 GAGGAGATGGAGAGAGGGCAAGATCAACAAGATCGGCCCCCGAGAAACCCCTACAACACC 846  
Db 781 GAGGAGATGGAGAGAGGGCAAGATCAACAAGATCGGCCCCCGAGAAACCCCTACAACACC 840  
Qy 847 CCCGTGTTCCGCTCATCAAGAAGAGACAGACCAAGATGGCGCAAGCTGGTGGAATTCGCG 906  
Db 841 CCCGTGTTCCGCTCATCAAGAAGAGAGACAGACCAAGATGGCGCAAGCTGGTGGAATTCGCG 900  
Qy 907 GAGCTGAACAGCGCACCAGAGCTTCTGGGAGGTGACGCTGGGCAATCCGCCACCCCGCC 966  
Db 901 GAGCTGAACAGCGCACCAGAGCTTCTGGGAGGTGACGCTGGGCAATCCGCCACCCCGCC 960  
Qy 967 GGCCTGAAGAGAGAGAGCGTGACCGTGTGGACGTGGCGACCGCTACTTTCAGCGTG 1026  
Db 961 GGCCTGAAGAGAGAGAGCGTGACCGTGTGGACGTGGCGACCGCTACTTTCAGCGTG 1020  
Qy 1027 CCCCTGACAGGAGACTTCCGCAAGTACACCGCTTTCACATCCCGAGATCAACAACGAG 1086  
Db 1021 CCCCTGACAGGAGACTTCCGCAAGTACACCGCTTTCACATCCCGAGATCAACAACGAG 1080  
Qy 1087 ACCCCCGGCATCCGCTACAGTAAAGTGTGCTGCCCGAGGCTGGAGGGCAGCCCGAC 1146  
Db 1081 ACCCCCGGCATCCGCTACAGTAAAGTGTGCTGCCCGAGGCTGGAGGGCAGCCCGAC 1140  
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGAGAGCCTTCCGCGCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGAGAGCCTTCCGCGCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACAGGCCCGCTGTAAGTGGGAGAGCACTGGAGATCGGCGACACCGCGCC 1266  
Db 1201 GTGATCTACAGGCCCGCTGTAAGTGGGAGAGCACTGGAGATCGGCGACACCGCGCC 1260  
Qy 1267 AAGATCAGGAGCTGCGCAAGCACTGCTGGTGGCTGGGGCTTACACACCCCGCAAGAG 1326  
Db 1261 AAGATCAGGAGCTGCGCAAGCACTGCTGGTGGCTGGGGCTTACACACCCCGCAAGAG 1320  
Qy 1327 CACCAGAGAGCGCCCTTCTGATGATGGCTACGAGCTGCAACCCCGCAAGAGTGGACC 1386  
Db 1321 CACCAGAGAGCGCCCTTCTGATGATGGCTACGAGCTGCAACCCCGCAAGAGTGGACC 1374  
Qy 1387 GTGAGCCCATCGAGCTGCCGAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTG 1446  
Db 1375 GTGAGCCCATCGAGCTGCCGAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTG 1434  
Qy 1447 GTGGGCAAGCTGAATGGGCGCAGCATCTACCCCGCATCAAGGTGCGCGACCTGTC 1506  
Db 1435 GTGGGCAAGCTGAATGGGCGCAGCATCTACCCCGCATCAAGGTGCGCGACCTGTC 1494  
Qy 1507 AAGCTGTGCGCGCGCCAGGCGCTGACGACATCTGCGCCCTGACCGAGAGGCGCGAG 1566  
Db 1495 AAGCTGTGCGCGCGCGCCAGGCGCTGACGACATCTGCGCCCTGACCGAGAGGCGCGAG 1554  
Qy 1567 CTGAGCTGGCCGAGAAACCGGAGATCTGCGGAGCGCCGTCGACGCGGCTGTACTAGAC 1626  
Db 1555 CTGAGCTGGCCGAGAAACCGGAGATCTGCGGAGCGCCGTCGACGCGGCTGTACTAGAC 1614  
Qy 1627 CCAGAGAGGACCTGTGGCGCGAGATCCAGAGAGGCGCGACGACGATGACCTACAG 1686  
Db 1615 CCAGAGAGGACCTGTGGCGCGAGATCCAGAGAGGCGCGACGACGATGACCTACAG 1674  
Qy 1687 ATCTTACAGAGCGCTTCAAGAACTGAAGACCGCAAGTACGCGCAAGATGCGCACCGCC 1746  
Db 1675 ATCTTACAGAGCGCTTCAAGAACTGAAGACCGCAAGTACGCGCAAGATGCGCACCGCC 1734  
Qy 1747 CACACCAACGCTGAAGAGCTGACCGAGGCGCTGCAGAGATCGCCATGGAGAGATC 1806  
Db 1735 CACACCAACGCTGAAGAGCTGACCGAGGCGCTGCAGAGATCGCCATGGAGAGATC 1794  
Qy 1807 GTGATCTGGGCGAAGACCCCAAGTTCCGCTGCGCCATCCAGAGAGGACCTGGGAGACC 1866

Db 1795 GTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAAAGGAGACCTGGAGACC 1854  
Qy 1867 TGGTGGACCGACTACTGGCAGGGCCACCTGGATCCCCCGAGTGGGAGTTTGTGAACACCCCC 1926  
Db 1855 TGGTGGACCGACTACTGGCAGGGCCACCTGGATCCCCCGAGTGGGAGTTTGTGAACACCCCC 1914  
Qy 1927 CCCCTGTTGAAGCTGTGGTACCAGCTGGAGAGAGGCCCATCATCGGCGCCGAGACCTTC 1986  
Db 1915 CCCCTGTTGAAGCTGTGGTACCAGCTGGAGAGAGGCCCATCATCGGCGCCGAGACCTTC 1974  
Qy 1987 TACGTGAGACGCGCGCCCAACCGGAGACCAAGATCGCAAGGGCGGTACTGTACCCGAC 2046  
Db 1975 TACGTGAGACGCGCGCCCAACCGGAGACCAAGATCGCAAGGGCGGTACTGTACCCGAC 2034  
Qy 2047 CGGGCGCGCAGAGATCTGTAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2106  
Db 2035 CGGGCGCGCAGAGATCTGTAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2094  
Qy 2107 GCCATCAGCTGGCCCTGACGAGCAGCGGACGAGGTGAACATCTGTACCGACGACCGAG 2166  
Db 2095 GCCATCAGCTGGCCCTGACGAGCAGCGGACGAGGTGAACATCTGTACCGACGACCGAG 2154  
Qy 2167 TAGCCCTTGGGCAATCTCCAGGCCAGCCGACAAAGAGCGAGAGCTGCTGAACCGAG 2226  
Db 2155 TAGCCCTTGGGCAATCTCCAGGCCAGCCGACAAAGAGCGAGAGCTGCTGAACCGAG 2214  
Qy 2227 ATCATCGAGCAGCTGATCAAGAGAGGAGAGGTGTACCTGAGCTGGTGCCCGCCACAG 2286  
Db 2215 ATCATCGAGCAGCTGATCAAGAGAGGAGAGGTGTACCTGAGCTGGTGCCCGCCACAG 2274  
Qy 2287 GGCATCGGCGGCAACGAGCAGATCGAAGCTGTGTAGCAAGGGCATCCGCAAGGTGCTG 2346  
Db 2275 GGCATCGGCGGCAACGAGCAGATCGAAGCTGTGTAGCAAGGGCATCCGCAAGGTGCTG 2334  
Qy 2347 TTCTCTGACGCGCATCGATCGCGCATCTGATCTACCAAGTACATGAGACGACCTGTACCTG 2406  
Db 2335 TTCTCTGACGCGCATCGATCGCGCATCTGATCTACCAAGTACATGAGACGACCTGTACCTG 2394  
Qy 2407 GGCAGCGCGCCCTTAGGATCGATTAAGCTTCCCGGGCTAGCACCGGT 2457  
Db 2395 GGCAGCGCGCCCTTAGGATCGATTAAGCTTCCCGGGCTAGCACCGGT 2445

## RESULT 10

US-10-190-435-9  
; Sequence 9, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBERG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190.435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 3930  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmut\_C  
US-10-190-435-9

Query Match 98.1%; Score 2415.6; DB 15; Length 3930;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2434; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 14 TGGCCGAGGCCATGAGCCAGGCCACCAGCCCAACATCTGTGATGACGCGAGCAACTTCA 73



Db	1487	TCGCGAGGCATGAGCAGGCCACGAGCGCAATCTCTGATGACGCGAGCAACTTCA	1546
Qy	74	AGGGCCCCAAGCGCATCATCAAGTGTCTTAACATCGCGCAAGAGGGGCCACATCGCCCGCA	133
Db	1547	AGGGCCCCAAGCGCATCATCAAGTGTCTTAACATCGCGCAAGAGGGGCCACATCGCCCGCA	1606
Qy	134	ACTGCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCGGCAAGAGGGGCCACAGATGA	193
Db	1607	ACTGCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCGGCAAGAGGGGCCACAGATGA	1666
Qy	194	AGGACTGACCGAGGGCCAGGCACCACTTCTCCGCGAGGACCTGGCCCTTCCCCCAGGGCA	253
Db	1667	AGGACTGACCGAGGGCCAGGCACCACTTCTCCGCGAGGACCTGGCCCTTCCCCCAGGGCA	1726
Qy	254	AGGCCCGGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCCACACAGCGCGGAGCTGC	313
Db	1727	AGGCCCGGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCCACACAGCGCGGAGCTGC	1786
Qy	314	AGGTGCGGCGCAACACCCCGCAGCGAGGCGCGCGCGCAGCGCCAGGGCAACCTTGAAC	373
Db	1787	AGGTGCGGCGCAACACCCCGCAGCGAGGCGCGCGCGCAGCGCCAGGGCAACCTTGAAC	1846
Qy	374	TCCCCCAGATCACCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTGGGCGGCGCAGATCA	433
Db	1847	TCCCCCAGATCACCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTGGGCGGCGCAGATCA	1906
Qy	434	AGGAGGCCCTGCTCGACACACCGGCCCGACGACACCGTCTGGAGGAGATGAGCCTGCCCG	493
Db	1907	AGGAGGCCCTGCTCGACACACCGGCCCGACGACACCGTCTGGAGGAGATGAGCCTGCCCG	1966
Qy	494	GCAAGTGGAGCCCAAGATGATCGCGCGCATTCGGCGCTTCATCAAGGTGGCGCAGTACG	553
Db	1967	GCAAGTGGAGCCCAAGATGATCGCGCGCATTCGGCGCTTCATCAAGGTGGCGCAGTACG	2026
Qy	554	ACAGATCTTGATCGAGATTCGCGCAAGAGGCCATCGGCAACCGTGTCTGATCGGCCCA	613
Db	2027	ACAGATCTTGATCGAGATTCGCGCAAGAGGCCATCGGCAACCGTGTCTGATCGGCCCA	2086
Qy	614	CCCCGTGACATCATCGCGCGCAACATGTCGACCGAGCTGGGTGACACCTGAACTTCC	673
Db	2087	CCCCGTGAAACATCATCGCGCGCAACATGTCGACCGAGCTGGGTGACACCTGAACTTCC	2146
Qy	674	CCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGAAAGCCCGGCATCGAGCGGCCCAAGG	733
Db	2147	CCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGAAAGCCCGGCATCGAGCGGCCCAAGG	2206
Qy	734	TGAAGCATGTGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA	793
Db	2207	TGAAGCATGTGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA	2266
Qy	794	TGGAGAGGAGGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTCAACACACCCCGTGT	853
Db	2267	TGGAGAGGAGGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTCAACACACCCCGTGT	2326
Qy	854	TCGCCATCAAGAAGAGCAGCACCAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA	913
Db	2327	TCGCCATCAAGAAGAGCAGCACCAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA	2386
Qy	914	ACAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCATCCCCCACCACCCCGCGGCTGA	973
Db	2387	ACAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCATCCCCCACCACCCCGCGGCTGA	2446
Qy	974	AGAAGAGAGAGCGTGCACCGTGTGAGCTGGGCGAGCGCTACTTACAGGCTGCCCTTG	1033
Db	2447	AGAAGAGAGAGCGTGCACCGTGTGAGCTGGGCGAGCGCTACTTACAGGCTGCCCTTG	2506
Qy	1034	ACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCGAGATCAACAGAGACCCCG	1093
Db	2507	ACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCGAGATCAACAGAGACCCCG	2566
Qy	1094	GCATCCGCTACCAGTACAAACGTGTGCTGCCCGCAGGCGTCGGAAGGGCAGCCCCAGCATCTTC	1153

2567	Db	GCATCCGCTACCAAGTACACGTCGTCGCCACAGGGCTGGAAAGGGCAGCCCCCAGCATCTTCC	2620
1154	Qy	AGAGCAGCATGACCAAGATCCTTGAGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCT	1213
2627	Db	AGAGCAGCATGACCAAGATCCTTGAGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCT	2686
1214	Qy	ACNAGGCCCTCTGTACGTGGGACAGCACTCTGAGATCGGCCAGACACCGGCCCAAGATCG	1273
2687	Db	ACCAGGCCCTCTGTACGTGGGACAGCACTCTGAGATCGGCCAGACACCGGCCCAAGATCG	2746
1274	Qy	AGGAGCTGCCAAGCACTCTCGCTGSGGCTTACCAACCCCGCAACAAGAACCCAGA	1333
2747	Db	AGGAGCTGCCAAGCACTCTCGCTGSGGCTTACCAACCCCGCAACAAGAACCCAGA	2806
1334	Qy	AGGAGCCCTCTCTGTGATGGGCTACAGAGCTGCACCCCGCAAAAGTGACCGTGCAGC	1393
2807	Db	AGGAGCCCTCTCTGTGATGGGCTACAGAGCTGCACCCCGCAAAAGTGACCGTGCAGC	2860
1394	Qy	CCATCGAGCTGCCCGAGAGAGGAGCTGGAACCTGGAACGACATCAGAGACTGGTGGCA	1453
2861	Db	CCATCGAGCTGCCCGAGAGAGGAGCTGGAACCTGGAACGACATCAGAGACTGGTGGCA	2920
1454	Qy	AGCTGAACCTGGGCGAGCCAGATCTACCCGGCATCAAGTGCSCAGCTGTCAAGCTGC	1513
2921	Db	AGCTGAACCTGGGCGAGCCAGATCTACCCGGCATCAAGTGCSCAGCTGTCAAGCTGC	2980
1514	Qy	TGCGCGCGCCAAAGGCCCTGACCAACATCGTGCCTCTGACCGAGGAGCCGAGCTGGAGC	1573
2981	Db	TGCGCGCGCCAAAGGCCCTGACCAACATCGTGCCTCTGACCGAGGAGCCGAGCTGGAGC	3040
1574	Qy	TGGCCGAGAACCCGAGATCCTGCGGAGCCCTGTGACGCGCGTGTACTAGACCCCAAGCA	1633
3041	Db	TGGCCGAGAACCCGAGATCCTGCGGAGCCCTGTGACGCGCGTGTACTAGACCCCAAGCA	3100
1634	Qy	AGGACCTGTGGCCGAGATCCAGACGAGGGCCACGACCAAGTGGACCTACCAAGCTTACC	1693
3101	Db	AGGACCTGTGGCCGAGATCCAGACGAGGGCCACGACCAAGTGGACCTACCAAGCTTACC	3160
1694	Qy	AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACACCA	1753
3161	Db	AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACACCA	3220
1754	Qy	ACGAGTGTAAGCAGCTGACACGAGCCGTGCAGAAAGATTCGCCATGGAGAGCATCGTGATCT	1813
3221	Db	ACGAGTGTAAGCAGCTGACACGAGCCGTGCAGAAAGATTCGCCATGGAGAGCATCGTGATCT	3280
1814	Qy	GGGGCAGACCCCAAGTTCCGCTCCCATCAGAGGAGGACCTGGGAGACCTGGTGGGA	1873
3281	Db	GGGGCAGACCCCAAGTTCCGCTCCCATCAGAGGAGGACCTGGGAGACCTGGTGGGA	3340
1874	Qy	CCGACTACTGGCAGGCCACTTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCCCTGG	1933
3341	Db	CCGACTACTGGCAGGCCACTTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCCCTGG	3400
1934	Qy	TGAAGCTGTGGTACCAAGCTGGAAGAGCCCATCATTCGCGCCGAGACCTTCTACGTGG	1993
3401	Db	TGAAGCTGTGGTACCAAGCTGGAAGAGCCCATCATTCGCGCCGAGACCTTCTACGTGG	3460
1994	Qy	ACGGCGCCGCCAACCCGGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC	2053
3461	Db	ACGGCGCCGCCAACCCGGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCC	3520
2054	Qy	GGCAGAGATCGGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCC	2113
3521	Db	GGCAGAGATCGGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCC	3580
2114	Qy	AGCTGGCCCTGCAGGCAGCGGAGGTGAACATCGTGACCGACAGCCAGCTACGCC	2173
3581	Db	AGCTGGCCCTGCAGGCAGCGGAGGTGAACATCGTGACCGACAGCCAGCTACGCC	3640
2174	Qy	TGGGCATCATCAGGCCAGCCCGACNAGAGCGAGAGCGAGCTGGTGAACCAAGATCATCG	2233
3641	Db	TGGGCATCATCAGGCCAGCCCGACNAGAGCGAGAGCGAGCTGGTGAACCAAGATCATCG	3700

```
Qy 2234 AGCAGCTGATCAAGAGGAGAGGTGTACTTGAGCTGGTCCCGCCACAAAGGCGATCG 2293
Db 3701 AGCAGCTGATCAAGAGGAGAGGTGTACTTGAGCTGGTCCCGCCACAAAGGCGATCG 3760
Qy 2294 GCGGCAACGACGATCGACAGCTGGTGACGAGGCGCATCCGCAAGGTGCTGTCTCGG 2353
Db 3761 GCGGCAACGACGATCGACAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCGG 3820
Qy 2354 ACGGCATCGATGGCGGCATCGTGATCTACCACTAGTACGAGCACTGTACGTGGGCGAG 2413
Db 3821 ACGGCATCGATGGCGGCATCGTGATCTACCACTAGTACGAGCACTGTACGTGGGCGAG 3880
Qy 2414 GCGGCGCTAGATCGATTTAAAGCTTCCCGGGGTAGCACCGGT 2457
Db 3881 GCGGCGCTAGATCGATTTAAAGCTTCCCGGGGTAGCACCGGT 3924

RESULT 11
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PFI8133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10

Query Match 98.0%; Score 2414; DB 15; Length 3930;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

Qy 14 TGGCGGAGGCATGAGCGGCGACGCGCCAGCCCAACATCCTGATGCGGCGAGCAACTTCA 73
Db 1487 TCGCGGAGGCATGAGCGGCGACGCGCCAGCCCAACATCCTGATGCGGCGAGCAACTTCA 1546
Qy 74 AGGCGCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGCGCAATCGCCGCA 133
Db 1547 AGGCGCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGCGCAATCGCCGCA 1606
Qy 134 ACTGCGCGCGCCCGCAAGAGGGCTGCTGGAAGTGCAGAGGAGGCGCACAGATGA 193
Db 1607 ACTGCGCGCGCCCGCAAGAGGGCTGCTGGAAGTGCAGAGGAGGCGCACAGATGA 1666
Qy 194 AGGACTGACCGAGCGCGCGCAACTTCTTCCGCGAGGACCTTGGCCCTCCCGAGGGCA 253
Db 1667 AGGACTGACCGAGCGCGCGCAACTTCTTCCGCGAGGACCTTGGCCCTCCCGAGGGCA 1726
Qy 254 AGGCGCGGAGTTCGCCAGCGAGCAGAAACCGCGCAACAGCCCAACGCGCGAGGCTGC 313
Db 1727 AGGCGCGGAGTTCGCCAGCGAGCAGAAACCGCGCAACAGCCCAACGCGCGAGGCTGC 1786
Qy 314 AGGTGCGGCGGCAACAACCCCGCAGCGAGGCGCGCGCGCGCGCGAGCGCACCTGAACT 373
Db 1787 AGGTGCGGCGGCAACAACCCCGCAGCGAGGCGCGCGCGCGCGCGAGCGCACCTGAACT 1846
Qy 374 TCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAGATCAAGGTGGCGGCGCAGATCA 433
Db 1846 TCCCGCAGATCACCTGTGTGGCAGCGCCCTGTGTGAGATCAAGGTGGCGGCGCAGATCA 433
```

```
Db 1847 TCCCGCAGATCACCTGTGTGGCAGCGCCCTGTGTGAGATCAAGGTGGCGGCGCAGATCA 1906
Qy 434 AGGAGGCGCTGTGTGGACACCGGCGCGCGACACACCGTGTGTGGAGAGATGAGCCTGCCCG 493
Db 1907 AGGAGGCGCTGTGTGGACTTCCGGCGCGCGACGACGACGCTGTGGAGAGATGAGCCTGCCCG 1966
Qy 494 GCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAAGGTGGCGCAGTACG 553
Db 1967 GCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAAGGTGGCGCAGTACG 2026
Qy 554 ACCAGATCTGTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGGCTGTGTGATCGGCGGCA 613
Db 2027 ACCAGATCTGTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGGCTGTGTGATCGGCGGCA 2086
Qy 614 CCCCCTGTAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGTGCAACCTGAACCTTCC 673
Db 2087 CCCCCTGTAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGTGCAACCTGAACCTTCC 2146
Qy 674 CCATCAGCCCATCGACACCGTGCCTGTAAGCTGAAGCCCGGCGCATGAGCGGCGGCGGCGG 733
Db 2147 CCATCAGCCCATCGACACCGTGCCTGTAAGCTGAAGCCCGGCGCATGAGCGGCGGCGGCGG 2206
Qy 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGA 2266
Qy 794 TGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCGCGGAGAACCCCTACAACAACCCCGTGT 853
Db 2267 TGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCGCGGAGAACCCCTACAACAACCCCGTGT 2326
Qy 854 TCGCCATCAAGAGAGAGGAGCAGACCAAGTGGCGCGCAAGCTGTGTGAGCTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAGAGAGGAGCAGACCAAGTGGCGCGCAAGCTGTGTGAGCTTCCGCGAGCTGA 2386
Qy 914 ACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCCACCCCGCGGCGCTGA 973
Db 2387 ACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCCACCCCGCGGCGCTGA 2446
Qy 974 AGAAGAGAGAGAGCGTGACCGTGTGACGCTGGGCGAGCGGCTACTTTCAGCGTGCCTCTGG 1033
Db 2447 AGAAGAGAGAGAGCGTGACCGTGTGACGCTGGGCGAGCGGCTACTTTCAGCGTGCCTCTGG 2506
Qy 1034 ACAGGAGCTTCCGCAAGTACACCGCTTTCACATCCCGAGGATCAACAACGAGACCCCGG 1093
Db 2507 ACAGGAGCTTCCGCAAGTACACCGCTTTCACATCCCGAGGATCAACAACGAGACCCCGG 2566
Qy 1094 GCATCCGCTACCAAGTCAAGCTGCTGCCAGAGGCTGGAAGGCGAGCCCGAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAAGTCAAGCTGCTGCCAGAGGCTGGAAGGCGAGCCCGAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCTTGGAGCCCTTTCGCGCGCGCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCTTGGAGCCCTTTCGCGCGCGCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACAGGCGCGCTGTAGTGGGCGAGCACTTGGAGATCGGCGAGCAGCGCGCGCAAGATCG 1273
Db 2687 ACCAGGCGCGCTGTAGTGGGCGAGCACTTGGAGATCGGCGAGCAGCGCGCGCAAGATCG 2746
Qy 1274 AGGAGCTGCGCAAGCAGCTGCTGCGTGGGCTTTCACCCCGCGCAAGAGCAGCAGCA 1333
Db 2747 AGGAGCTGCGCAAGCAGCTGCTGCGTGGGCTTTCACCCCGCGCAAGAGCAGCAGCA 2806
Qy 1334 AGGAGCGCGCTTTCCTGTGGATCGAGCTGCAACCCCGCGCAAGTGGAGCGGTGAGC 1393
Db 2807 AGGAGCGCGCTTTCCTGTGGATCGAGCTGCAACCCCGCGCAAGTGGAGCGGTGAGC 2860
Qy 1394 CCATCGAGCTGCGCGAGAGAGAGCTGGAACCGTGAACGAGATCCAGAGCTGTGTGGGCA 1453
Db 2861 CCATCGAGCTGCGCGAGAGAGAGCTGGAACCGTGAACGAGATCCAGAGCTGTGTGGGCA 2920
Qy 1454 AGCTGAACCTGGGCGAGCAGATCTACCCCGCGCATAGGTGCGGCGAGCTGTGAGCTGC 1513
Db 2921 AGCTGAACCTGGGCGAGCAGATCTACCCCGCGCATAGGTGCGGCGAGCTGTGAGCTGC 2980
```

Qy	1514	TGCGCGGCCCAAGGCCCTTGACCGACATCGTGCCCTTGACCGAGGAGGCCGAGCTGGAGC	1573
Db	2981	TGCGCGCGCCAGGCCCTGACCGACATCGTGCCCTTGACCGAGGAGGCCGAGCTGGAGC	3040
Qy	1574	TGGCCGAGAACCGCGAGATCTCTGCGAGAGCCCGTGACGCGCGTGTACTACGACCCAGCA	1633
Db	3041	TGGCCGAGAACCGCGAGATCTCTGCGAGAGCCCGTGACGCGCGTGTACTACGACCCAGCA	3100
Qy	1634	AGGACCTGGTGGCCGAGATCCAGAAGCAGAGGCCACGACCACTGAGACCTACAGATCTACC	1693
Db	3101	AGGACCTGGTGGCCGAGATCCAGAAGCAGAGGCCACGACCACTGAGACCTACAGATCTACC	3160
Qy	1694	AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCACCGCCACACCA	1753
Db	3161	AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCACCGCCACACCA	3220
Qy	1754	ACGACGTGAAGACGCTGACCGAGGCCGTGCAGAGATCGCCATGGAGAGCATCGTGATCT	1813
Db	3221	ACGACGTGAAGACGCTGACCGAGGCCGTGCAGAGATCGCCATGGAGAGCATCGTGATCT	3280
Qy	1814	GGGCGAAGACCCCAAGTTCCGCCCTGCCCATCCAGAAGAGAGACCTGGGAGACCTGGTGA	1873
Db	3281	GGGCGAAGACCCCAAGTTCCGCCCTGCCCATCCAGAAGAGAGACCTGGGAGACCTGGTGA	3340
Qy	1874	CCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTTCGTGAACACACCCCCCTCG	1933
Db	3341	CCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTTCGTGAACACACCCCCCTCG	3400
Qy	1934	TGAAGCTGTGGTACCAAGCTGGAGAGAGGCCCATCATCGCGCGCGAGACCTTCTACGTGG	1993
Db	3401	TGAAGCTGTGGTACCAAGCTGGAGAGAGGCCCATCATCGCGCGCGAGACCTTCTACGTGG	3460
Qy	1994	ACGCGCGCGCAACCGCGAGACCAAGATTCGGCAAGCGCGGCTACGTACCGACCGGGGCC	2053
Db	3461	ACGCGCGCGCAACCGCGAGACCAAGATTCGGCAAGCGCGGCTACGTACCGACCGGGGCC	3520
Qy	2054	GGCAGAGATCGTGAGCCTGCACGAGACCAACACAGAGAGCCGAGCTCGAGGCCATCC	2113
Db	3521	GGCAGAGATCGTGAGCCTGCACGAGACCAACACAGAGAGCCGAGCTCGAGGCCATCC	3580
Qy	2114	AGCTGGCCCTTCAGGACAGCGGCGAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC	2173
Db	3581	AGCTGGCCCTTCAGGACAGCGGCGAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC	3640
Qy	2174	TGGGCATCATCCAGGCCAGCCGACAAGAGCGAGAGCGAGTGGTGAACCGAGATCATCG	2233
Db	3641	TGGGCATCATCCAGGCCAGCCGACAAGAGCGAGAGCGAGTGGTGAACCGAGATCATCG	3700
Qy	2234	AGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGTGCCGCCCAACAGGGCATCG	2293
Db	3701	AGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGTGCCGCCCAACAGGGCATCG	3760
Qy	2294	CGGSCAACGAGCAGATCGACAAGCTGTGTAGCAAGGGCATCCGCAAGTGCTGTCTCTCG	2353
Db	3761	CGGSCAACGAGCAGATCGACAAGCTGTGTAGCAAGGGCATCCGCAAGTGCTGTCTCTCG	3820
Qy	2354	ACGGCATCGATGGCGGCATCGTGATCTTACAGTATCATGGAAGCGACCTGTACGTGGGAGCG	2413
Db	3821	ACGGCATCGATGGCGGCATCGTGATCTTACAGTATCATGGAAGCGACCTGTACGTGGGAGCG	3880
Qy	2414	CGCGCCCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT	2457
Db	3881	CGCGCCCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT	3924

RESULT 12

US-10-190-435-11

US-10-130-433-11  
; Sequence 11, Application US/10190435

; Publication No. US20030143248A1

; GENERAL INFORMATION:

APPLICANT: ZUR MEGEDE, Jan

APPLICANT: BARNETT, Susan W.

```

: APPLICANT: LIAN, Ying
: APPLICANT: ENGELBRECHT, Susan
: APPLICANT: VAN RENSBURG, Estrelita J.
: TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
: TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
: FILE REFERENCE: PP18133.003 / 2302-18133
: CURRENT APPLICATION NUMBER: US/10/190,435
: CURRENT FILING DATE: 2002-12-30
: NUMBER OF SEQ ID NOS: 319
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 11
: LENGTH: 3930
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: GagCmplPolmutIna_C
: US-10-190-435-11

```

Query Match	98.0%	Score 2414	DB 15	Length 3930
Best Local Similarity	99.5%	Pred. No. 0		
Matches 2433	Conservative	0	Mismatches	5
			Indels	6
			Gaps	1
Qy	14	TGCGCAGGCGCATGAGCCAGCGCCACACAGCGCCAAACATCTCTGATGCGAGCGAGCAACTTCA	73	
Db	1487	TCGCGAGGCCATGAGCCAGCGCCACAGCGCCAAACATCTCTGATGCGAGCGAGCACTTCA	1546	
Qy	74	AGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGCGCAAGGAGGCCACATCGCCCGCA	133	
Db	1547	AGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGCGCAAGGAGGCCACATCGCCCGCA	1606	
Qy	134	ACTGCGCGCGCCCGCCAGAGAGGGCTGCTGGAAGTGGCGACGAGGAGGGCCACAGATGA	193	
Db	1607	ACTGCGCGCGCCCGCCAGAGAGGGCTGCTGGAAGTGGCGCAAGGAGGGCCACAGATGA	1666	
Qy	194	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGGGCTTCTCCCGCAGGCA	253	
Db	1667	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGGGCTTCTCCCGCAGGCA	1726	
Qy	254	AGGCGCGCAGTTTCCCGAGCGAGCAGAAACGCGGCCAAACAGCCCAACAGCGCGAGCTGC	313	
Db	1727	AGGCGCGCAGTTTCCCGAGCGAGCAGAAACGCGGCCAAACAGCCCAACAGCGCGAGCTGC	1786	
Qy	314	AGGTGCGCGCGACAACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGCCAGGGCACTTGAAT	373	
Db	1787	AGGTGCGCGCGACAACCCCGCAGCGAGCGCGCGCCGAGGCCAGGGCACTTGAAT	1846	
Qy	374	TCGCCCAGATCACCTGTGTGCGACGCGCCCTGTGTGAGCATCAAGGTGGCGGCCAGATCA	433	
Db	1847	TCGCCCAGATCACCTGTGTGCGACGCGCCCTGTGTGAGCATCAAGGTGGCGGCCAGATCA	1906	
Qy	434	AGGAGGCTGTGTGACACCGCGCGCGACGACACGTTGTGAGGAGATGAGCCTGCCG	493	
Db	1907	AGGAGGCTGTGTGGCCACCGCGCGCGACGACACGTTGTGAGGAGATGAGCCTGCCG	1966	
Qy	494	GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGCGCCAGTACG	553	
Db	1967	GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGCGCCAGTACG	2026	
Qy	554	ACCAGATCTCTGATCGAGATCTCGCGCAAGAGGCCATCGCACCTGCTGATCGGCCCA	613	
Db	2027	ACCAGATCTCTGATCGAGATCTCGCGCAAGAGGCCATCGCACCTGCTGATCGGCCCA	2086	
Qy	614	CCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTTGAACTTCC	673	
Db	2087	CCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTTGAACTTCC	2146	
Qy	674	CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGCGCCCAAGG	733	
Db	2147	CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGCGCCCAAGG	2206	
Qy	734	TGAAGCAGTGGCCCTCTGAACCGAGGAGAGATCAAGGCGCTTGAACCGGCATCTGCGAGAGA	793	
Db	2207	TGAAGCAGTGGCCCTCTGAACCGAGGAGAGATCAAGGCGCTTGAACCGGCATCTGCGAGAGA	2266	



QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGCAAGGAGGCGCCACATCGCCCGCA 133  
 DB |||||  
 2801 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGCAAGGAGGCGCCACATCGCCCGCA 2860  
 QY 134 ACTGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTGCAGGCGCAAGGAGGCGCCACAGATGA 193  
 DB |||||  
 2861 ACTGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTGCAGGCGCAAGGAGGCGCCACAGATGA 2920  
 QY 194 AGGACTGCAACCGAGCGCGAGGCCAACTTCTTCGCGAGGACCTGGGCTTCCCGCAGGGCA 253  
 DB |||||  
 2921 AGGACTGCAACCGAGCGCGAGGCCAACTTCTTCGCGAGGACCTGGGCTTCCCGCAGGGCA 2980  
 QY 254 AGGCGCGCGAGTTCGCCAGCGAGCAGAAACCGCGCCCAACAGCCCAACAGCGCGAGCTGC 313  
 DB |||||  
 2981 AGGCGCGCGAGTTCGCCAGCGAGCAGAAACCGCGCCCAACAGCCCAACAGCGCGAGCTGC 3040  
 QY 314 AGGTGCGCGCGAGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 DB |||||  
 3041 AGGTGCGCGCGAGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100  
 QY 374 TCCCCCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433  
 DB |||||  
 3101 TCCCCCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3160  
 QY 434 AGGAGGCGCTGTGAGACACCGGCGCGAGCAGACACCGTGTGGAGGAGATGAGCGTCCCGG 493  
 DB |||||  
 3161 AGGAGGCGCTGTGAGCAGCGGCGCGAGCAGACACCGTGTGGAGGAGATGAGCGTCCCGG 3220  
 QY 494 GCAAGTGGAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCGAGTACG 553  
 DB |||||  
 3221 GCAAGTGGAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAGGTGCGCGAGTACG 3280  
 QY 554 ACCAGATCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATCGGCGCCCA 613  
 DB |||||  
 3281 ACCAGATCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATCGGCGCCCA 3340  
 QY 614 CCCCCTGGAACATCATCTGCGCGCAACATGTCAGACCCAGCTGGGCTGCACCTTGAATCTCC 673  
 DB |||||  
 3341 CCCCCTGGAACATCATCTGCGCGCGCAACATGTCAGACCCAGCTGGGCTGCACCTTGAATCTCC 3400  
 QY 674 CCATCAGCCCCATCGAGACCGTGTGCGCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGG 733  
 DB |||||  
 3401 CCATCAGCCCCATCGAGACCGTGTGCGCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGG 3460  
 QY 734 TGNAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793  
 DB |||||  
 3461 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520  
 QY 794 TGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCCCGAGAAACCCCTTACAACACCCCGTGT 853  
 DB |||||  
 3521 TGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCCCGAGAAACCCCTTACAACACCCCGTGT 3580  
 QY 854 TCGCCATCAAGAAGAGGAGCAGCAACCAAGTGGCGCAAGCTGTGTGACTTCGCGAGCTGA 913  
 DB |||||  
 3581 TCGCCATCAAGAAGAGGAGCAGCAACCAAGTGGCGCAAGCTGTGTGACTTCGCGAGCTGA 3640  
 QY 914 ACNAGCGCACCCAGGACTTCTGGGAGGTGCGAGCTGGGCGATCCCCCAACCGCGCGGCTGA 973  
 DB |||||  
 3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGCGAGCTGGGCGATCCCCCAACCGCGCGGCTGA 3700  
 QY 974 AGAAGAAGAGGCGTGAACCGTGTGAGCGTGGGCGACGCTACTTTCAGCGTGGCCCCCTGG 1033  
 DB |||||  
 3701 AGAAGAAGAGGCGTGAACCGTGTGAGCGTGGGCGACGCTACTTTCAGCGTGGCCCCCTGG 3760  
 QY 1034 ACAGGACTTTCGCAAGTACACCGCTTTCACATCCCGCAGCATCAACAGAGACCCCGG 1093  
 DB |||||  
 3761 ACAGGACTTTCGCAAGTACACCGCTTTCACATCCCGCAGCATCAACAGAGACCCCGG 3820  
 QY 1094 GCATCCGCTACCGATCAACAGTGTCTGCGCGAGGCTGGAGGCGAGCCCGCAGCATCTTCC 1153  
 DB |||||  
 3821 GCATCCGCTACCGATCAACAGTGTCTGCGCGAGGCTGGAGGCGAGCCCGCAGCATCTTCC 3880

QY 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGCGCGCAACCCCGAGATCGTGTATCT 1213  
 DB |||||  
 3881 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGCGCGCAACCCCGAGATCGTGTATCT 3940  
 QY 1214 ACAGGCCCCCTCTGTACTGTGGGAGCGACTGTGGAGATCGGCGCAGCAACCGGCGCAAGATCG 1273  
 DB |||||  
 3941 ACAGGCCCCCTCTGTACTGTGGGAGCGACTGTGGAGATCGGCGCAGCAACCGGCGCAAGATCG 4000  
 QY 1274 AGGAGTGGCGCAAGCACTGTCTGCGTGGGCTTACCAACCCCGCAACAAGAGCACCA 1333  
 DB |||||  
 4001 AGGAGTGGCGCAAGCACTGTCTGCGTGGGCTTACCAACCCCGCAACAAGAGCACCA 4060  
 QY 1334 AGGAGCCCCCTTCTCTGTGGATGAGGTGACAGCTGCACCCCGCAACAAGTGGACCGTGCAGC 1393  
 DB |||||  
 4061 AGGAGCCCCCTTCTCTGCCAT-----CGAGCTGCACCCCGACAAAGTGGACCGTGCAGC 4114  
 QY 1394 CCATCGAGTGTCCCGAGAGGAGCTGGACCGTGAACGACATCCAGAGACTGGTGGGCA 1453  
 DB |||||  
 4115 CCATCGAGTGTCCCGAGAGAGGAGCTGGACCGTGAACGACATCCAGAGACTGGTGGGCA 4174  
 QY 1454 AGCTGAATGSGCGCAGCCAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGTGC 1513  
 DB |||||  
 4175 AGCTGAATGSGCGCAGCCAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGTGC 4234  
 QY 1514 TCGCGCGCGCCAAAGGCTTGACCGACATCTGTGCCCTGACCGAGGAGGCGAGCTGGAGC 1573  
 DB |||||  
 4235 TCGCGCGCGCCAAAGGCTTGACCGACATCTGTGCCCTGACCGAGGAGGCGAGCTGGAGC 4294  
 QY 1574 TGGCCGAGAAACCGCGAGATCTCTGCGGAGCCCGTGCACGGCGTGTACTAGACCCAGCA 1633  
 DB |||||  
 4295 TGGCCGAGAAACCGCGAGATCTCTGCGGAGCCCGTGCACGGCGTGTACTAGACCCAGCA 4354  
 QY 1634 AGGACTGTGTGCGCGAGATCCAGAGCAGGCGCACGACAGTGGACTTACCAGATCTACC 1693  
 DB |||||  
 4355 AGGACTGTGTGCGCGAGATCCAGAGCAGGCGCACGACAGTGGACTTACCAGATCTACC 4414  
 QY 1694 AGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCAACCA 1753  
 DB |||||  
 4415 AGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCAACCA 4474  
 QY 1754 AGA CGTGAAGCAGCTGACCGGAGCGCTGCAGAGATTCGCCATGAGAGGATCGTGTATCT 1813  
 DB |||||  
 4475 ACGACGTGAAGCAGCTGACCGGAGCGCTGCAGAGATTCGCCATGAGAGGATCGTGTATCT 4534  
 QY 1814 GGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTCGGAGACCTCGTGGGA 1873  
 DB |||||  
 4535 GGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTCGGAGACCTCGTGGGA 4594  
 QY 1874 CCGACTACTGGCAGGCGCACCTCGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGG 1933  
 DB |||||  
 4595 CCGACTACTGGCAGGCGCACCTCGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGG 4654  
 QY 1934 TGAAGCTGTGTACCGAGTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTAGCTGG 1993  
 DB |||||  
 4655 TGAAGCTGTGTACCGAGTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTAGCTGG 4714  
 QY 1994 ACGGCGCCCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCCGGGCC 2053  
 DB |||||  
 4715 ACGGCGCCCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCCGGGCC 4774  
 QY 2054 GGCAGAGATCTGTGAGCTTGACCGAGACCAACCAAGAGACCGAGCTGCGAGGCGCATCC 2113  
 DB |||||  
 4775 GGCAGAGATCTGTGAGCTTGACCGAGACCAACCAAGAGACCGAGCTGCGAGGCGCATCC 4834  
 QY 2114 AGCTGGCCCTTGAGGACGCGGAGCGAGGTGAAACATCTGTGACCGAGCAGCAGTACGCC 2173  
 DB |||||  
 4835 AGCTGGCCCTTGAGGACGCGGAGCGAGGTGAAACATCTGTGACCGAGCAGCAGTACGCC 4894  
 QY 2174 TGGGCATCATCAGGCGCCAGCGCAAGAGCGAGAGCGAGCTGGTGAACACAGATCATCG 2233  
 DB |||||  
 4895 TGGGCATCATCAGGCGCCAGCGCAAGAGCGAGAGCGAGCTGGTGAACACAGATCATCG 4954  
 QY 2234 AGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCGCCCAAGGGCATCG 2293

```
Db 4955 AGCAGCTGATCAAGAAAGAGAGGTGACCTGAGCTGGGTGCCCGCCCAACAAGGGCATCG 5014
Qy 2294 GCGCAACAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTCTGG 2353
Db 5015 GCGCAACAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTCTGG 5074
Qy 2354 ACGGCATCGATGGCGGCATCGTGTACTACAGTATACATGGACGACCTGTACGTGGGACGG 2413
Db 5075 ACGGCATCGATGGCGGCATCGTGTACTACAGTATACATGGACGACCTGTACGTGGGACGG 5134
Qy 2414 GCGGCCCTAGCATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db 5135 GCGGCCCTAGCATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 5178

RESULT 14
US-10-190-305A-82
; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190.305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 82
; TYPE: DNA
; LENGTH: 5184
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: TatRevNefgagPolina C
US-10-190-305A-82

Query Match 98.08; Score 2414; DB 16; Length 5184;
Best Local Similarity 99.58; Pred. No. 0;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

Qy 14 TGGCCGAGGCGCATGAGCCAGGCCACCAAGCGCCCAACTCTGTATCGCAGCGCAGCAACTTCA 73
Db 2741 TCGCCGAGGCGCATGAGCCAGGCCACCAAGCGCCCAACTCTGTATCGCAGCGCAGCAACTTCA 2800
Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 133
Db 2801 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATCGCCCGCA 2860
Qy 134 ACTGCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCAGGAGGGCCACAGATGA 193
Db 2861 ACTGCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCAGGAGGGCCACAGATGA 2920
Qy 194 AGGACTGCAACGAGCGCGCAGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCCCAGGGCA 253
Db 2921 AGGACTGCAACGAGCGCGCAGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCCCAGGGCA 2980
Qy 254 AGGCCCGCGAGTTCCCCAGCGAGAGAAACCGCGCCAAACAGCCCAACAGCCCGAGCTGC 313
Db 2981 AGGCCCGCGAGTTCCCCAGCGAGAGAAACCGCGCCAAACAGCCCAACAGCCCGAGCTGC 3040
Qy 314 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
Db 3041 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100
Qy 374 TCCCCCAGATCACCTGTGTGACGGCCCTCTGTGTAGCATCAAGGTGGGGCGCCAGATCA 433
Db 3101 TCCCCCAGATCACCTGTGTGACGGCCCTCTGTGTAGCATCAAGGTGGGGCGCCAGATCA 3160
Qy 434 AGGAGGCCCTGTGTGGAACACCGGCGCGGACGACACCGGTGTGTGGAGGAGATGAGCCTGCCG 493
```

```
Db 3161 AGGAGGCCCTGTGTGGAACACCGGCGCGGACGACACCGTGTGTGGAGGAGATGAGCCTGCCG 3220
Qy 494 GCAAGTGAAGCCCAAGATGATCGGCGGCATTCGGCGGCTTTCATCAAGGTGCGCCAGTACG 553
Db 3221 GCAAGTGAAGCCCAAGATGATCGGCGGCATTCGGCGGCTTTCATCAAGGTGCGCCAGTACG 3280
Qy 554 ACCAGATCTGTATCGAGATCTGCGGGCAAGAGCCATCGGCACCCGTGTGATCGGCCCA 613
Db 3281 ACCAGATCTGTATCGAGATCTGCGGGCAAGAGCCATCGGCACCCGTGTGATCGGCCCA 3340
Qy 614 CCCCCTGTAACATCATCGGCGGCAACATGCTGACCCAGCTGGGTGACCCCTGAACCTTCC 673
Db 3341 CCCCCTGTAACATCATCGGCGGCAACATGCTGACCCAGCTGGGTGACCCCTGAACCTTCC 3400
Qy 674 CCATCAGCCCCCATCGAGACCGTGTCCGTGAAGCTGAAAGCCCGGATGGAAGCCGCCAAGG 733
Db 3401 CCATCAGCCCCCATCGAGACCGTGTCCGTGAAGCTGAAAGCCCGGATGGAAGCCGCCAAGG 3460
Qy 734 TGAAGCAGTGGCCCTTGACCGAGAGAAAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGA 793
Db 3461 TGAAGCAGTGGCCCTTGACCGAGAGAAAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGA 3520
Qy 794 TGGAGAGAGGGGCAAGATCAACCAAGATCGGCGCGGAGAACCCCTACAACACCCCTGT 853
Db 3521 TGGAGAGAGGGGCAAGATCAACCAAGATCGGCGCGGAGAACCCCTACAACACCCCTGT 3580
Qy 854 TCGCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913
Db 3581 TCGCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCCACCCCGCGGCTGA 973
Db 3641 ACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCCACCCCGCGGCTGA 3700
Qy 974 AGAAGAAAGAGGGTGACCCGTCTGSAAGTGGGCGAGGCTTACTTACAGGTGCCCCCTGG 1033
Db 3701 AGAAGAAAGAGGGTGACCCGTCTGSAAGTGGGCGAGGCTTACTTACAGGTGCCCCCTGG 3760
Qy 1034 ACCAGGACTTTCGCAAGTACACCGCTTACACCTCCCGAGCATCAACAGAGACCCCG 1093
Db 3761 ACCAGGACTTTCGCAAGTACACCGCTTACACCTCCCGAGCATCAACAGAGACCCCG 3820
Qy 1094 GCATCCGCTACCAAGTACAAAGTGTCTGCCAGGGCTGGAAGGGCAGGCCACAGCATCTTCC 1153
Db 3821 GCATCCGCTACCAAGTACAAAGTGTCTGCCAGGGCTGGAAGGGCAGGCCACAGCATCTTCC 3880
Qy 1154 AGACGACATGACCAAGATCTTGGAGGCCCTTCCGCGCGCGCAACCCCGAGATCGTGATCT 1213
Db 3881 AGACGACATGACCAAGATCTTGGAGGCCCTTCCGCGCGCGCAACCCCGAGATCGTGATCT 3940
Qy 1214 ACCAGGCCCTTGTACGTGGGCGAGCGACCTGGAGATCGGCCAGCACCGGCCAAGATCG 1273
Db 3941 ACCAGGCCCTTGTACGTGGGCGAGCGACCTGGAGATCGGCCAGCACCGGCCAAGATCG 4000
Qy 1274 AGGAGCTGCGCAAGCAAGCCTGCTCGCTGGGGCTTTCACACCCCGCAAGAGCACAGA 1333
Db 4001 AGGAGCTGCGCAAGCAAGCCTGCTCGCTGGGGCTTTCACACCCCGCAAGAGCACAGA 4060
Qy 1334 AGGAGGCCCTTCTCTGTGTGATGGGCTACGAGCTGACCCCGGACCAAGTGAACCGTGAGC 1393
Db 4061 AGGAGGCCCTTCTCTGTGTGATGGGCTACGAGCTGACCCCGGACCAAGTGAACCGTGAGC 4114
Qy 1394 CCATCAGAGCTGCCCGAGAGAGGAGCTGGAGCCGTGAAACGACATCCAGAAGCTGGTGGCA 1453
Db 4115 CCATCAGAGCTGCCCGAGAGAGGAGCTGGAGCCGTGAAACGACATCCAGAAGCTGGTGGCA 4174
Qy 1454 AGCTGAACCTGGGCGAGCCAGATCTACCCCGGCACTCAAGGTGCGCCAGCTGTGCAAGCTGC 1513
Db 4175 AGCTGAACCTGGGCGAGCCAGATCTACCCCGGCACTCAAGGTGCGCCAGCTGTGCAAGCTGC 4234
Qy 1514 TGGCGGCGCGCAAGGGCCCTGACCGACATCTGTGCCCTTGACCGGAGGAGCGGAGCTGAGC 1573
```



```
Db 4235 TCGCGGGCCCAAGGCGCTTGACCGACATCGTGTGCCCTGACCGAGGAGGCGGAGCTGGAGC 4294
QY 1574 TGGCCGAGAACCGCGAGATCTTGGCGAGCGCGTGCACGGCGTGTAAGGACCCAGCA 1633
Db 4295 TGGCCGAGAACCGCGAGATCTTGGCGAGCGCGTGCACGGCGTGTAAGGACCCAGCA 4354
QY 1634 AGGACCTGTGTGGCCGAGATCTCAGAAGCAGGGCCACGACAGTGGACTTACCAGATCTACC 1693
Db 4355 AGGACCTGTGTGGCCGAGATCTCAGAAGCAGGGCCACGACAGTGGACTTACCAGATCTACC 4414
QY 1694 AGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGACCGGCCACACCA 1753
Db 4415 AGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGACCGGCCACACCA 4474
QY 1754 ACGAGCTGAAGCAGCTGACCGAGGCGCGTGCAGAGATCGCCATGGAGAGATCGTGATCT 1813
Db 4475 ACGAGCTGAAGCAGCTGACCGAGGCGCGTGCAGAGATCGCCATGGAGAGATCGTGATCT 4534
QY 1814 GGGGCAAGACCCCAAGTTCCGCTCGCCCATCCAGAAGGAGACCTGGGAGACTGGTGGA 1873
Db 4535 GGGGCAAGACCCCAAGTTCCGCTCGCCCATCCAGAAGGAGACCTGGGAGACTGGTGGA 4594
QY 1874 CCGACTACTGGCAGGCGCACCTGATCCCCGAGTGGGAGTTCTGTGAACACCCCCCTGG 1933
Db 4595 CCGACTACTGGCAGGCGCACCTGATCCCCGAGTGGGAGTTCTGTGAACACCCCCCTGG 4654
QY 1934 TGAAGCTGTGTACAGCTGGGAGAGGAGCCCATCATCGGCGCGAGACTTCTAGCTGG 1993
Db 4655 TGAAGCTGTGTGTACAGCTGGGAGAGGAGCCCATCATCGGCGCGAGACTTCTAGCTGG 4714
QY 1994 ACGGCGCCGCCAAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTGACCGACCGGGGCC 2053
Db 4715 ACGGCGCCGCCAAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTGACCGACCGGGGCC 4774
QY 2054 GGCAGAGATCTGTAGCTTGA CCGAGACCAACCAAGAGACCGAGCTGCGAGGCCATCC 2113
Db 4775 GGCAGAGATCTGTAGCTTGA CCGAGACCAACCAAGAGACCGAGCTGCGAGGCCATCC 4834
QY 2114 AGCTGGCCCTGCAGGACGCGCAGGAGTGAACATCGTGACCGACGCGCAGTACGCC 2173
Db 4835 AGCTGGCCCTGCAGGACGCGCAGGAGTGAACATCGTGACCGACGCGCAGTACGCC 4894
QY 2174 TGGGCATCATCAGGCGCCAGCGCCGACAAGAGCGAGCGAGCTGTGTGAACCCAGATCATCG 2233
Db 4895 TGGGCATCATCAGGCGCCAGCGCCGACAAGAGCGAGCGAGCTGTGTGAACCCAGATCATCG 4954
QY 2234 AGCAGCTGATCAAGAGGAGAGGTGTA CTTAGCTGGGTGCGGCCCAAGAGGCGATCG 2293
Db 4955 AGCAGCTGATCAAGAGGAGAGGTGTA CTTAGCTGGGTGCGGCCCAAGAGGCGATCG 5014
QY 2294 GCGGCAACAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGACAGGTGCTGTCTCTGG 2353
Db 5015 GCGGCAACAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCTGG 5074
QY 2354 ACGGCATCGATGCGGCGCATCTGTATCTTACAGTACATGACGACCTGTACGTGGGCGCG 2413
Db 5075 ACGGCATCGATGCGGCGCATCTGTATCTTACAGTACATGACGACCTGTACGTGGGCGCG 5134
QY 2414 GCGGCGCTTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 2457
Db 5135 GCGGCGCTTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 15  
US-10-190-435-13  
; Sequence 13, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEYER, Jan W.  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.

```
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 3531  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_c  
US-10-190-435-13  
  
Query Match 96.8%; Score 2383.6; DB 15; Length 3531;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 2414; Conservative 0; Mismatches 24; Indels 6; Gaps 1;  
  
QY 14 TGGCGAGGCCATGAGCGCCACCCAGCGCCAAACATCTCTGATGCGAGCGCAGCAACTTCA 73  
Db 1088 TGGCGAGGCCATGAGCGCCACCCAGCGCCAAACATCTCTGATGCGAGAGCAACTTTAAA 1147  
QY 74 AGGGCCCCAAGCGCATCATCAAGTCTTCAACTCTGCGGCAAGGAGGCCACATGCCCCGCA 133  
Db 1148 AGGGCCCCAAGCGCATCATCAAGTCTTCAACTCTGCGGCAAGGAGGCCACATGCCCCGCA 1207  
QY 134 ACTGCGCGCCCCCGCAGAGAGGGCTGTGGAAGTGGCGCAAGGAGGCCACATGCCCCGCA 193  
Db 1208 ACTGCGCGCCCCCGCAGAGAGGGCTGTGGAAGTGGCGCAAGGAGGCCACATGCCCCGCA 1267  
QY 194 AGGACTGCAACCGAGCGCCAGGCGCAACTTCTTCCGCGAGGACCTTGGCTTCCCGCAGGGCA 253  
Db 1268 AGGACTGCAACCGAGCGCCAGGCGCAACTTCTTCCGCGAGGACCTTGGCTTCCCGCAGGGCA 1327  
QY 254 AGGCGCGCGAGTTCCCGAGCGAGCAACCGCGCCAAACAGCCCAACAGCGCGGAGCTGC 313  
Db 1328 AGGCGCGCGAGTTCCCGAGCGAGCAACCGCGCCAAACAGCCCAACAGCGCGGAGCTGC 1387  
QY 314 AGGTGCGCGCGACAAACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGCGCCACCTGAACT 373  
Db 1388 AGGTGCGCGCGACAAACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGCGCCACCTGAACT 1447  
QY 374 TCCCCAGATCACCTGTGCGAGCGCCCTCTGTGAGCATCAAGGTGGGCGCCAGATCA 433  
Db 1448 TCCCCAGATCACCTGTGCGAGCGCCCTCTGTGAGCATCAAGGTGGGCGCCAGATCA 1507  
QY 434 AGGAGCCCTGTGGAACCGCGCGCGACGACACCGTGTGAGGAGATGAGCTGCCCCG 493  
Db 1508 AGGAGCCCTGTGGAACCGCGCGCGACGACACCGTGTGAGGAGATGAGCTGCCCCG 1567  
QY 494 GCAAGTGAAGCCCAAGATGATCGGCGCATCGGCGCTTCAAGGTGCGCGAGTACG 553  
Db 1568 GCAAGTGAAGCCCAAGATGATGCGCGCATCGGCGCTTCAAGGTGCGCGAGTACG 1627  
QY 554 ACCAGATCTGTGAGATCTGCGCAAGAGGCCATCGGCACCGCTGCTGATCGGCCCCA 613  
Db 1628 ACCAGATCTGTGAGATCTGCGCAAGAGGCCATCGGCACCGCTGCTGATCGGCCCCA 1687  
QY 614 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTGAACTTCC 673  
Db 1688 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTGAACTTCC 1747  
QY 674 CCATAGCCCCCATCGAGAGCGTCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGG 733  
Db 1748 CCATAGCCCCCATCGAGAGCGTCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGG 1807  
QY 734 TGAAGCAGTGGGCCCTTGACCGCGAGGAGAGATCAAGGCCCTTGACCGGCATCTGCGAGGAGA 793  
Db 1808 TGAAGCAGTGGGCCCTTGACCGCGAGGAGAGATCAAGGCCCTTGACCGGCATCTGCGAGGAGA 1867  
QY 794 TGAAGAGGAGGCGAAGATCACCAAGATCGGCCCCGAGAACCCCTTCAACACCCCGTGT 853  
|||||
```

Db 1868 TGGAGAGGAGGGCAAGATCACCAAGATCGGGCCCGGAGAACCCCTACAAACACCCCGGTGT 1927  
Qy 854 TCGCCATCAAGAAAGAGACAGCAACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913  
Db 1928 TCGCCATCAAGAAAGAGACAGCAACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 1987  
Qy 914 ACAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCAATCCCCACCCCGCGGCTGA 973  
Db 1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCAATCCCCACCCCGCGGCTGA 2047  
Qy 974 AGAAGAAAGAGAGCGTGAACCGTGTGGACGTGGGCGAGCGCTACTTTCAGCGTGGCCCTGG 1033  
Db 2048 AGAAGAAAGAGAGCGTGAACCGTGTGGACGTGGGCGAGCGCTACTTTCAGCGTGGCCCTGG 2107  
Qy 1034 ACAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAGACCCCG 1093  
Db 2108 ACAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAGACCCCG 2167  
Qy 1094 GCATCCGCTACCAAGTACAAAGTGTGCTGCCCCAGGCTGGAAGGCGAGCCCCAGCATCTTCC 1153  
Db 2168 GCATCCGCTACCAAGTACAAAGTGTGCTGCCCCAGGCTGGAAGGCGAGCCCCAGCATCTTCC 2227  
Qy 1154 AGACGACATGACCAAGATCCTGGAGCCCTTTCGCGCCGCAACCCCGAGATCGTGATCT 1213  
Db 2228 AGACGACATGACCAAGATCCTGGAGCCCTTTCGCGCCGCAACCCCGAGATCGTGATCT 2287  
Qy 1214 ACCAGGCCCTCTGTACGTGGGCGAGCAGCTGGAGATCGGCCAGCACCGGCCCAAGATCG 1273  
Db 2288 ACCAGGCCCTCTGTACGTGGGCGAGCAGCTGGAGATCGGCCAGCACCGGCCCAAGATCG 2347  
Qy 1274 AGGAGCTGGCGAGCAGCTGCTGGCTGGGCTTTCACCAACCCCGACAGAGACACAGA 1333  
Db 2348 AGGAGCTGGCGAGCAGCTGCTGGCTGGGCTTTCACCAACCCCGACAGAGACACAGA 2407  
Qy 1334 AGGAGCCCCCTTCTGTGATGGCTACGAGCTGCACCCGACAAAGTGGACCGTGCAGC 1393  
Db 2408 AGGAGCCCCCTTCTGTGATGGCTACGAGCTGCACCCGACAAAGTGGACCGTGCAGC 2461  
Qy 1394 CCATCGAGCTGCCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGCA 1453  
Db 2462 CCATCGAGCTGCCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGCA 2521  
Qy 1454 AGCTGAATGGGCGAGCAGATCTACCCCGGCATCAAGGTGCGGCAGCTGTGCAAGCTGC 1513  
Db 2522 AGCTGAATGGGCGAGCAGATCTACCCCGGCATCAAGGTGCGGCAGCTGTGCAAGCTGC 2581  
Qy 1514 TGGCGGCGCCAAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAGCGGAGCTGGAGC 1573  
Db 2582 TGGCGGCGCCAAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAGCGGAGCTGGAGC 2641  
Qy 1574 TGGCGGAGAACCGGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCACAGCA 1633  
Db 2642 TGGCGGAGAACCGGAGATCCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCACAGCA 2701  
Qy 1634 AGGACCTGGTGGCGGAGATCCAGAAAGAGGCGCACGACAGTGGACCTTACAGATCTACC 1693  
Db 2702 AGGACCTGGTGGCGGAGATCCAGAAAGAGGCGCACGACAGTGGACCTTACAGATCTACC 2761  
Qy 1694 AGGAGCCCTTTCAGAACTGAAAGACCGGAGTACGCAAGATGCGGACCGCCACACCA 1753  
Db 2762 AGGAGCCCTTTCAGAACTGAAAGACCGGAGTACGCAAGATGCGGACCGCCACACCA 2821  
Qy 1754 ACAGCTGAAGAGAGCTGACCGAGGCCGTGAGAGAGATCGGCATGCGAGAGCATCGTGATCT 1813  
Db 2822 ACAGCTGAAGAGAGCTGACCGAGGCCGTGAGAGAGATCGGCATGCGAGAGCATCGTGATCT 2881  
Qy 1814 GGGCAAGACCCCAAGTTCCGCTGCCCCATCCAGAAAGGAGACCTTGGGAGACCTTGGTGA 1873  
Db 2882 GGGCAAGACCCCAAGTTCCGCTGCCCCATCCAGAAAGGAGACCTTGGGAGACCTTGGTGA 2941  
Qy 1874 CCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGAAACACCCCGCCCTGG 1933  
Db 2942 CCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGAAACACCCCGCCCTGG 3001

Qy 1934 TGAAGCTGTGTATCAGCTGGAGAAAGAGGCCCATCATCGGCGCCGAGACCTTCTACGTGG 1993  
Db 3002 TGAAGCTGTGTATCAGCTGGAGAAAGAGGCCCATCATCGGCGCCGAGACCTTCTACGTGG 3061  
Qy 1994 ACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCTAGTGAGACGACCGGGGCC 2053  
Db 3062 ACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCTAGTGAGACGACCGGGGCC 3121  
Qy 2054 GGCAGAAAGATCGTGAGCTGACCGAGACCAACCAAGAAAGCGAGCTGCAGGGCCATCC 2113  
Db 3122 GGCAGAAAGATCGTGAGCTGACCGAGACCAACCAAGAAAGCGAGCTGCAGGGCCATCC 3181  
Qy 2114 AGCTGGCCCTGCGAGACAGCGGCGAGCGGTGAACATGTGAACCGACAGCCAGTACGCC 2173  
Db 3182 AGCTGGCCCTGCGAGACAGCGGCGAGCGGTGAACATGTGAACCGACAGCCAGTACGCC 3241  
Qy 2174 TGGGCATCATCCAGGCCCGCCGACAGAGCGAGGAGGAGCTGGTGAACCGAGATCATCG 2233  
Db 3242 TGGGCATCATCCAGGCCCGCCGACAGAGCGAGGAGGAGCTGGTGAACCGAGATCATCG 3301  
Qy 2234 AGCAGCTGATCAAGAAAGAGAGGTGTACTGTGGTCCCGCCACCAAGGGCATCG 2293  
Db 3302 AGCAGCTGATCAAGAAAGAGAGGTGTACTGTGGTCCCGCCACCAAGGGCATCG 3361  
Qy 2294 GCGCAACGAGCAGATCGACAAGCTGGTGAACAAAGGGCATCCGCAAGGTGCTGTTCCTGG 2353  
Db 3362 GCGCAACGAGCAGATCGACAAGCTGGTGAACAAAGGGCATCCGCAAGGTGCTGTTCCTGG 3421  
Qy 2354 ACGGCATCGATGGCGGCGATCGTGATCTACCGTATACGAGCAGCTGTACGTTGGCAGG 2413  
Db 3422 ACGGCATCGATGGCGGCGATCGTGATCTACCGTATACGAGCAGCTGTACGTTGGCAGG 3481  
Qy 2414 GCGGCCCTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 2457  
Db 3482 GCGGCCCTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 3525

Search completed: June 2, 2005, 04:23:40  
Job time : 1419.33 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 03:50:55 ; Search time 10513 Seconds  
(without alignments)  
11324.506 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457  
Sequence: 1 gtcagcaccaccatggccga.....gggctagcaccgtgaattc 2457

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	100.0	2457	6	AX455916
2	2436.2	99.2	2463	6	AX455915
3	2415.4	98.3	2469	6	AX455914
4	2040	83.0	2300	6	BD263705
5	2040	83.0	2300	6	CQ870575
6	2040	83.0	2300	6	AR373388
7	2019.2	82.2	2306	6	BD263704
8	2019.2	82.2	2306	6	CQ870574
9	2019.2	82.2	2306	6	AR373387
10	2000.8	81.4	9165	6	AX427930
11	1998.4	81.3	2312	6	BD263706
12	1998.4	81.3	2312	6	CQ870576
13	1998.4	81.3	2312	6	AR373389
14	1992.4	81.1	9788	6	AX427936
15	1978.4	80.5	9169	6	AX427931
16	1975	80.4	9194	6	AX427926
17	1973.4	80.3	9194	6	AX427925
18	1973.4	80.3	12411	6	AX427927
19	1970	80.2	9785	6	AX427938

20	1966.6	80.0	9167	6	AX427933	AX427933 Sequence
21	1966.6	80.0	9170	6	AX427928	AX427928 Sequence
22	1966.6	80.0	9782	6	AX427935	AX427935 Sequence
23	1966.6	80.0	9783	6	AX427934	AX427934 Sequence
24	1966.6	80.0	9792	6	AX427932	AX427932 Sequence
25	1965	80.0	9189	6	AX427921	AX427921 Sequence
26	1962	79.9	4353	12	AF287353	AF287353 Synthetic
27	1945.4	79.2	3015	6	AX455946	AX455946 Sequence
28	1928.8	78.5	3009	6	AX455987	AX455987 Sequence
29	1915.8	78.0	4319	6	BD263639	BD263639 Improved
30	1915.8	78.0	4319	6	CQ870498	CQ870498 Sequence
31	1915.8	78.0	4319	6	AR373322	AR373322 Sequence
32	1905.2	77.5	4352	12	AF287352	AF287352 Synthetic
33	1878.6	76.5	1965	6	AX455952	AX455952 Sequence
34	1872.8	76.2	2299	6	BD263703	BD263703 Improved
35	1872.8	76.2	2299	6	CQ870573	CQ870573 Sequence
36	1872.8	76.2	2299	6	AR373386	AR373386 Sequence
37	1859.8	75.7	2577	6	AX457088	AX457088 Sequence
38	1852.2	75.4	4343	6	AX306428	AX306428 Sequence
39	1852	75.4	2305	6	BD263702	BD263702 Improved
40	1852	75.4	2305	6	CQ870572	CQ870572 Sequence
41	1852	75.4	2305	6	AR373385	AR373385 Sequence
42	1848.2	75.2	4341	6	AX306429	AX306429 Sequence
43	1838.6	74.8	1977	6	AX455954	AX455954 Sequence
44	1825.6	74.3	7897	6	AX427923	AX427923 Sequence
45	1772	72.1	3024	12	AF287355	AF287355 Synthetic

## ALIGNMENTS

RESULT 1  
AX455916  
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 32 from Patent WO0204493.  
ACCESSION AX455916  
VERSION AX455916.1 GI:21714901  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 32 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES  
Location/Qualifiers  
source  
1. .2457  
/organism="synthetic construct"  
/mol type="unassigned DNA"  
/db xref="taxon:32630"  
/note="PR975YMW"

## ORIGIN

Query Match 100.0%; Score 2457; DB 6; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 1.5e-247;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GTCAGCCCACTTCAAGGCGCCCAAGCGCATCATCAAGTCTCAACTCGCGCAAGGCGC 120  
Db 1 GTCAGCCCACTTCAAGGCGCCCAAGCGCATCATCAAGTCTCAACTCGCGCAAGGCGC 120  
Qy 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTCTCAACTCGCGCAAGGCGC 120  
Db 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTCTCAACTCGCGCAAGGCGC 120  
Qy 121 CACATCGCCCGCAACTCGCCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCCGCAACTCGCCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Qy 181 GGCCACCAAGTAGAGGACTGCACCGAGCGCCAGCCCAACTTCTTCGCGAGGAGCCTGGCC 240

181	Db	GGCCACACAGATGAAGGACTGTCACCGAGCGCCAGGSCCAACTTCTTCTTCGCGAGGACCTGTGGCC	240
241	Qy	TTTCCCCAGGGCAAGGCCCGCGAGTTCTCCACGAGCAGAGAAACCGCGCAACACAGCGCCCAACC	300
241	Db	TTTCCCCAGGGCAAGGCCCGCGAGTTCTCCACGAGCAGAGAAACCGCGCAACACAGCGCCCAACC	300
301	Qy	AGCCGCGAGCTGACAGTGCGCGGCGCAAAACCCCGCAGCAGGCGCGCGCCGAGCGCCAG	360
301	Db	AGCCGCGAGCTGACAGTGCGCGGCGCAAAACCCCGCAGCAGGCGCGCGCCGAGCGCCAG	360
361	Qy	GGCACCCCTGAATTTCCCCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCATCAGGTG	420
361	Db	GGCACCCCTGAATTTCCCCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG	420
421	Qy	GGCGGCCAGATCAAGGAGGCCCTGCTCGACAACCGCGGCCGACGACACCGTGTCTGGAGGAG	480
421	Db	GGCGGCCAGATCAAGGAGGCCCTGCTCGACAACCGCGGCCGACGACACCGTGTCTGGAGGAG	480
481	Qy	ATGAGCCTGCCC CGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAG	540
481	Db	ATGAGCCTGCCC CGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAG	540
541	Qy	GTGCGCCAGTACGACACAGATCTGTGATCTGCGGCGAAGAGGCCATCGGCAACCGTG	600
541	Db	GTGCGCCAGTACGACACAGATCTGTGATCTGCGGCGAAGAGGCCATCGGCAACCGTG	600
601	Qy	CTGATCGGCCCCACCCCGCTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
601	Db	CTGATCGGCCCCACCCCGCTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
661	Qy	ACCTCTGAATTTCCCCCATCAGCCCCATCGAGACCGTGCCTGTGAAGCTGAAAGCCCGGCATG	720
661	Db	ACCTCTGAATTTCCCCCATCAGCCCCATCGAGACCGTGCCTGTGAAGCTGAAAGCCCGGCATG	720
721	Qy	GACGGCCCCAAGGTGAAGCTGTGCCCTGTGACCGAGGAGAAGATCAAGGCCCTTGACCGCC	780
721	Db	GACGGCCCCAAGGTGAAGCTGTGCCCTGTGACCGAGGAGAAGATCAAGGCCCTTGACCGCC	780
781	Qy	ATCTGCGAGGAGATGGAGAGGGGGCAAGATCACCAAGATCGSCCCCGAGAACCCCTAC	840
781	Db	ATCTGCGAGGAGATGGAGAGGGGGCAAGATCACCAAGATCGSCCCCGAGAACCCCTAC	840
841	Qy	AACACCCCCGTTTCGCCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAC	900
841	Db	AACACCCCCGTTTCGCCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAC	900
901	Qy	TTTCGCGAGCTGAACAAGCGCACCCAGGACTTTCTGGGAGTGCAGCTGGGGCATCCCCAC	960
901	Db	TTTCGCGAGCTGAACAAGCGCACCCAGGACTTTCTGGGAGTGCAGCTGGGGCATCCCCAC	960
961	Qy	CCCCCGGCTGAGAGAGAGAGAGCGTGCAGCTGTGACGCTGGGCGAGCGCTACTTC	1020
961	Db	CCCCCGGCTGAGAGAGAGAGAGCGTGCAGCTGTGACGCTGGGCGAGCGCTACTTC	1020
1021	Qy	AGCGTGCCTTCGACGAGGACTTCCGCAAGTACACCCGCTTCACCATCCCGAGCATCAAC	1080
1021	Db	AGCGTGCCTTCGACGAGGACTTCCGCAAGTACACCCGCTTCACCATCCCGAGCATCAAC	1080
1081	Qy	AACGAGACCCCGGCATCCGCTACCACTGTAACAGTGTCTGACGCTGGAGCGCGCAAC	1140
1081	Db	AACGAGACCCCGGCATCCGCTACCACTGTAACAGTGTCTGACGCTGGAGCGCGCAAC	1140
1141	Qy	CCGAGCATCTTCGAGGAGCATGACCAAGATCTCTGAGCGCTTCGCGGCCCGCAACCCC	1200
1141	Db	CCGAGCATCTTCGAGGAGCATGACCAAGATCTCTGAGCGCTTCGCGGCCCGCAACCCC	1200
1201	Qy	GAGATCGTGATCTACGAGGCCCCCTCTGTACTGTGGGCGAGCACTTGGAGATCGGCGAGCAC	1260
1201	Db	GAGATCGTGATCTACGAGGCCCCCTCTGTACTGTGGGCGAGCACTTGGAGATCGGCGAGCAC	1260
1261	Qy	CGCGCCCAAGATCGAGGAGCTGCGCAAGCACTGTCTGCTGGGCTTTCACCAACCCCGCAC	1320
1261	Db	CGCGCCCAAGATCGAGGAGCTGCGCAAGCACTGTCTGCTGGGCTTTCACCAACCCCGCAC	1320

QY	1321	AAGAACACACGAAGAGAGCCCCCTTCTCTGCCCATCTGAGCTGCACCCCGACAAGTGGACC	1381
DB	1321	AAGAAAGCACAGGAAGAGAGCCCCCTTCTCTGCCCATCTGAGCTGCACCCCGACAAGTGGACC	1380
QY	1381	GTGCAGGCCCATCTCAGACTGCGCGAGAGGAGAGCTGGACCGTGAACGACATCTCAGAAGCTG	1440
DB	1381	GTGCAGGCCCATCTCAGACTGCGCGAGAGGAGAGCTGGAACCGTGAACGACATCTCAGAAGCTG	1440
QY	1441	GTGGCGAAGCTGAACTGGGCGACGACAGATCTAACC CGGCATCAAGGTGCGCCAGCTGTGC	1500
DB	1441	GTGGCGAAGCTGAACTGGGCGACGACAGATCTAACC CGGCATCAAGGTGCGCCAGCTGTGC	1500
QY	1501	AAGCTGCTGCGCGCGCCCAAGGCCCTTGACCGAGCATCTGTGCCCTGTGACCCGAGGAGGCCGAG	1560
DB	1501	AAGCTGCTGCGCGCGCCCAAGGCCCTTGACCGAGCATCTGTGCCCTGTGACCCGAGGAGGCCGAG	1560
QY	1561	CTGGAGCTGGCCGAGAAACCGCGAGATCTCTGCGGAGCCCGTGCACGCGCTGTACTACGAC	1620
DB	1561	CTGGAGCTGGCCGAGAAACCGCGAGATCTCTGCGGAGCCCGTGCACGCGCTGTACTACGAC	1620
QY	1621	CCCAGCAAGAGACTGTGTGGCCGAGATCCAGAGCGAGGCCACGACCMGTGGACTTACCAG	1680
DB	1621	CCCAGCAAGAGACTGTGTGGCCGAGATCCAGAGCGAGGCCACGACCMGTGGACTTACCAG	1680
QY	1681	ATCTTACCAAGAGCCCTTCAAGAACTCTCAAGACCGGCAAGTACGCAAGATGCGCACCGCC	1740
DB	1681	ATCTTACCAAGAGCCCTTCAAGAACTCTCAAGACCGGCAAGTACGCAAGATGCGCACCGCC	1740
QY	1741	CACACCAACGAGCTGAAGCAGCTGACCGAGGCCGTGCAGAAAGTACGCCATGGAGAGCATC	1800
DB	1741	CACACCAACGAGCTGAAGCAGCTGACCGAGGCCGTGCAGAAAGTACGCCATGGAGAGCATC	1800
QY	1801	GTGATCTGGGGCAGAGCCCAAGTTCCGCTCCCATCCAGAAAGGAGACTCTGGAGACC	1860
DB	1801	GTGATCTGGGGCAGAGCCCAAGTTCCGCTCCCATCCAGAAAGGAGACTCTGGAGACC	1860
QY	1861	TGGTGGACCCGACTACTGCGAGGCCACCTGGATCCCCGAGTGGGAGTTTCGTGAACACCCCC	1920
DB	1861	TGGTGGACCCGACTACTGCGAGGCCACCTGGATCCCCGAGTGGGAGTTTCGTGAACACCCCC	1920
QY	1921	CCCTGTGTGAAGCTGTGGTACCAGCTGGAGAGAGGAGCCCATCATCGCGCCGAGAGACTTC	1980
DB	1921	CCCTGTGTGAAGCTGTGGTACCAGCTGGAGAGAGGAGCCCATCATCGCGCCGAGAGACTTC	1980
QY	1981	TACGTGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGCCGCTACGTGACCGAC	2040
DB	1981	TACGTGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGCCGCTACGTGACCGAC	2040
QY	2041	CGGGCCCGGCAAGAGATCGTGAGCTGACCGAGACCAACCGAGAGAGAGAGAGAGAGAGAG	2100
DB	2041	CGGGCCCGGCAAGAGATCGTGAGCTGACCGAGACCAACCGAGAGAGAGAGAGAGAGAGAG	2100
QY	2101	GCGATCCAGCTGGGCCCTTGAGGACGCGGAGAGAGAGTGAACATCGTGACCGGAGCGAG	2160
DB	2101	GCGATCCAGCTGGGCCCTTGAGGACGCGGAGAGAGAGTGAACATCGTGACCGGAGCGAG	2160
QY	2161	TACGCCCTTGGGCATCATCCAGGCCCGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2220
DB	2161	TACGCCCTTGGGCATCATCCAGGCCCGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2220
QY	2221	ATCATCGAGCAGCTGTATCAAGAGGAGAGAGTGTACTGTAGCTGGGTGCCCGCCACAG	2280
DB	2221	ATCATCGAGCAGCTGTATCAAGAGGAGAGAGTGTACTGTAGCTGGGTGCCCGCCACAG	2280
QY	2281	GCGATCGCGGCAACGAGCAGATCGACAGCTGGTGTAGCAAGGAGAGAGAGAGAGAGAG	2340
DB	2281	GCGATCGCGGCAACGAGCAGATCGACAGCTGGTGTAGCAAGGAGAGAGAGAGAGAGAG	2340
QY	2341	TTCCTGAGCGGCATCGATGGCGGCATCTGTGATCTTACCAAGTACATGAGAGAGAGAGAG	2400
DB	2341	TTCCTGAGCGGCATCGATGGCGGCATCTGTGATCTTACCAAGTACATGAGAGAGAGAGAG	2400





```
|||||
1681 TACCAGATCTACCGAGGCCCTTCAAGAACCTTGAAGACGGCAAGTACGCCAAGATGCGC 1740
QY ACCGCCCAACCAACGACGTAAGACAGTGCACGAGGGCGTGCAGAAATGCCATGGAG 1794
Db ACCGCCCAACCAACGACGTAAGACAGTGCACGAGGGCGTGCAGAAATGCCATGGAG 1800
QY AGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGG 1854
Db AGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGG 1860
QY GAGACCTGGTGGACCGACTACTTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTGAAC 1914
Db GAGACCTGGTGGACCGACTACTTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTGAAC 1920
QY ACCCCCCCTGGTGAAGCTGTGTACAGCTGTGAGAGAGAGCCCATCATCGGCGCCGAG 1974
Db ACCCCCCCTGGTGAAGCTGTGTGTACAGCTGTGAGAGAGAGCCCATCATCGGCGCCGAG 1980
QY ACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGGCTACGTG 2034
Db ACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGGCTACGTG 2040
QY ACCGACCGGGCGCGCAGAAATCGTGAGCCTGACCGAGACCAACCAAGAGACCGAG 2094
Db ACCGACCGGGCGCGCAGAAATCGTGAGCCTGACCGAGACCAACCAAGAGACCGAG 2100
QY CTGCAGGCCATCCAGCTGGCCCTGAGAGACAGCGGAGGAGGTGAACATCGTGACCGAC 2154
Db CTGCAGGCCATCCAGCTGGCCCTGAGAGACAGCGGAGGAGGTGAACATCGTGACCGAC 2160
QY AGCCAGTAGCCCTGGGCATCATCAGGCCCGACCGGAGGAGGAGCGAGCTGGTG 2214
Db AGCCAGTAGCCCTGGGCATCATCAGGCCCGACCGGAGGAGGAGCGAGCTGGTG 2220
QY AACCCAGATCATCAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGTGGTGCCCGCC 2274
Db AACCCAGATCATCAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGTGGTGCCCGCC 2280
QY CACAAGGGGATCGGGCGCAACAGAGCATGCAAGAGTGGTGAGCAAGGGGATCGCGAAG 2334
Db CACAAGGGGATCGGGCGCAACAGAGCATGCAAGAGTGGTGAGCAAGGGGATCGCGAAG 2340
QY GTGCTGTTCCTGGAGCGGATCGATCGGCGCATCGTACCTACCTACCTACCTACCTAC 2394
Db GTGCTGTTCCTGGAGCGGATCGATCGGCGCATCGTACCTACCTACCTACCTACCTAC 2400
QY TACGTGGGCGAGCGCGGCGCTAGGATCGATTAAGAGTTCCTGGGCTAGCACCGGTGAA 2454
Db TACGTGGGCGAGCGCGGCGCTAGGATCGATTAAGAGTTCCTGGGCTAGCACCGGTGAA 2460
QY TTC 2457
Db TTC 2463
```

```
RESULT 3
AX455914
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 30 from Patent WO020493.
ACCESSION AX455914
VERSION AX455914.1 GI:21714899
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE
1 zur Mercede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL polypeptides and uses thereof
Patent: WO 0204493-A 30 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
Location/Qualifiers
```

```
source
1. 2469
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975 (+)"
ORIGIN
Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
Best Local Similarity 99.3%; Pred. No. 3.2e-243;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;
QY 1 GTCAGCGCCACATGCGCGAGGCCCATGAGCGCAGGCGCACAGCGCCCAATCCTGTATGCAG 60
Db 1 GTCAGCGCCACCATGCGCGAGGCCCATGAGCGCAGGCGCACAGCGCCCAATCCTGTATGCAG 60
QY 61 GCGAGCAACTTCAAGGGCCCCAAGGGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120
Db 61 GCGAGCAACTTCAAGGGCCCCAAGGGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120
QY 121 CACATCGCCCCCAACTGCGCGCCCGCGAGAGGGCTGCTGGAAGTGGGCAAGGAG 180
Db 121 CACATCGCCCCCAACTGCGCGCCCGCGAGAGGGCTGCTGGAAGTGGGCAAGGAG 180
QY 181 GCGCACCAAGATGAAGGACTGCAACGAGCGCCAGGCCAACTTCTTCCGCGAGACCTGGCC 240
Db 181 GCGCACCAAGATGAAGGACTGCAACGAGCGCCAGGCCAACTTCTTCCGCGAGACCTGGCC 240
QY 241 TTCCCGCAGGGGAGAGGCCCGGAGTTCCCGCAGGAGGAGAACCGCGCCCAACAGCCCCACC 300
Db 241 TTCCCGCAGGGGAGAGGCCCGGAGTTCCCGCAGGAGGAGAACCGCGCCCAACAGCCCCACC 300
QY 301 AGCCCGAGCTGCGAGGTGCGCGGAGCAAAACCCGCGAGGCGCGCGCGAGCGCCAG 360
Db 301 AGCCCGAGCTGCGAGGTGCGCGGAGCAAAACCCGCGAGGCGCGCGCGAGCGCCAG 360
QY 361 GCGCACCTGAACTTCCCGCAGATCACTGCGGAGCGCCCTTGGTGGAGCATCAAGGTG 420
Db 361 GCGCACCTGAACTTCCCGCAGATCACTGCGGAGCGCCCTTGGTGGAGCATCAAGGTG 420
QY 421 GCGCGCCAGATCAAGAGGGCGCTGCTGAGCAACCGCGCGCGAGCACCGTGTGAGGAG 480
Db 421 GCGCGCCAGATCAAGAGGGCGCTGCTGAGCAACCGCGCGCGAGCACCGTGTGAGGAG 480
QY 481 ATGAGCCTGCGCGGAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCAATCAAG 540
Db 481 ATGAGCCTGCGCGGAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCAATCAAG 540
QY 541 GTGCGCCAGTACGACCAAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCG 600
Db 541 GTGCGCCAGTACGACCAAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCG 600
QY 601 CTGATCGGCGCCACCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCGCCACCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
QY 661 ACCCTGAACTTCCCGCATCAGCCCCCATCGAGACCGCTGCGCGGAGTGAAGCCCGGCA 720
Db 661 ACCCTGAACTTCCCGCATCAGCCCCCATCGAGACCGCTGCGCGGAGTGAAGCCCGGCA 720
QY 721 GACGGCCCCAAGTGAAGCAGTGGGCCCTGACCGGAGGAGAGATCAAGGCGCTGACCGCC 780
Db 721 GACGGCCCCAAGTGAAGCAGTGGGCCCTGACCGGAGGAGAGATCAAGGCGCTGACCGCC 780
QY 781 ATCTCGGAGGAGTGGAGAGGGGCAAGATCACCAGATCGGCGCGGAGAACCCCTAC 840
Db 781 ATCTCGGAGGAGTGGAGAGGGGCAAGATCACCAGATCGGCGCGGAGAACCCCTAC 840
QY 841 AACACCCCGGTTCGCCATCAAGAGAGGAGACACCAAGTGGCGCAAGCTGTGTGAC 900
Db 841 AACACCCCGGTTCGCCATCAAGAGAGGAGACACCAAGTGGCGCAAGCTGTGTGAC 900
QY 901 TTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCAATCCCCAC 960
Db 901 TTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCAATCCCCAC 960
```

Db 901 TTCCGCGAGCTGAACACAGCGCACCCAGGACTTTCTGGGAGGTGCGAGCTGGGCAATCCCCCAC 960  
Qy 961 CCCGCGGCTGAAGAAAGAGAGCTGACCGTGTGGAGCGTGGGCGAGCGCTACTTTC 1020  
Db 961 CCCGCGGCTGAAGAAAGAGAGCTGACCGTGTGGAGCGTGGGCGAGCGCTACTTTC 1020  
Qy 1021 AGCGTGCCCTGGAGAGGACTTTCGGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGCCCTGGAGAGGACTTTCGGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGGCGATCCGCTACAGTACAACTGTGCTGCCCGGAGGCTGGAAGGCGAGC 1140  
Db 1081 AACGAGACCCCGGCGATCCGCTACAGTACAACTGTGCTGCCCGGAGGCTGGAAGGCGAGC 1140  
Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCGCGCCCGCAACCCC 1200  
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCGCGCCCGCAACCCC 1200  
Qy 1201 GAGATCGTGATCTACCA-----GGCCCCCTGTAGTGGGCGAGGACTTGGAGTCGGC 1254  
Db 1201 GAGATCGTGATCTACCAAGTACATGAGCGACCTGTACGTGGGCGAGGACTTGGAGTCGGC 1254  
Qy 1255 CAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGCGCTGGGCTTCAACACCC 1314  
Db 1261 CAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGCGCTGGGCTTCAACACCC 1320  
Qy 1315 CCCGACAAAGACACCAAGAGGAGCGCCCTTCTCTGCCCAT-----CGAGCTGCACCCC 1368  
Db 1321 CCCGACAAAGACACCAAGAGGAGCGCCCTTCTCTGCCCAT-----CGAGCTGCACCCC 1380  
Qy 1369 GACAAGTGAGCCGTGAGCCCATCGAGCTGCCCGGAGAGGAGCTGACCGTGAACGAC 1428  
Db 1381 GACAAGTGAGCCGTGAGCCCATCGAGCTGCCCGGAGAGGAGCTGACCGTGAACGAC 1440  
Qy 1429 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCGGAGGAGCTTACCCCGGATCAAGGTG 1488  
Db 1441 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCGGAGGAGCTTACCCCGGATCAAGGTG 1500  
Qy 1489 CGCCAGCTGTGCAAGCTGTCTCGGCGGCGCAAGGCGCTTACCGAGATCTGTCGCCCTGACC 1548  
Db 1501 CGCCAGCTGTGCAAGCTGTCTCGGCGGCGCAAGGCGCTTACCGAGATCTGTCGCCCTGACC 1560  
Qy 1549 GAGAGGCGGAGCTGTGGAGCTGGCGAGAACCGCGAGATCTGTGCGGAGCGCGTGCACGGC 1608  
Db 1561 GAGAGGCGGAGCTGTGGAGCTGGCGAGAACCGCGAGATCTGTGCGGAGCGCGTGCACGGC 1620  
Qy 1609 GTGTACTACGACCCGAGGAGCTGTGGCGGAGATCCAGAGCGAGGCGGAGGAGCAG 1668  
Db 1621 GTGTACTACGACCCGAGGAGCTGTGGCGGAGATCCAGAGCGAGGCGGAGGAGCAG 1680  
Qy 1669 TGGACCTACGAGATCTACCGAGGCGCTTCAAGAACCTGAGAGCGGCAAGTACGCCAAG 1728  
Db 1681 TGGACCTACGAGATCTACCGAGGCGCTTCAAGAACCTGAGAGCGGCAAGTACGCCAAG 1740  
Qy 1729 ATGGCGACCGGCGGACCAACAGAGCTGAAGAGCTGACCGAGGCGGCTGCAAGAGATCGCC 1788  
Db 1741 ATGGCGACCGGCGGACCAACAGAGCTGAAGAGCTGACCGAGGCGGCTGCAAGAGATCGCC 1800  
Qy 1789 ATGAGAGCATCTGTATCTGGGGGAGACCCCGCAAGTTCGCGCTGCCCATCCAGAGGAG 1848  
Db 1801 ATGAGAGCATCTGTATCTGGGGGAGACCCCGCAAGTTCGCGCTGCCCATCCAGAGGAG 1860  
Qy 1849 ACCTGGGAGACTGTGTGAGCGGACTTACTTGGGAGGCGGAGGAGGAGGAGGAGTTC 1908  
Db 1861 ACCTGGGAGACTGTGTGAGCGGACTTACTTGGGAGGCGGAGGAGGAGGAGGAGTTC 1920  
Qy 1909 GTGAACACCCCGGCTGTGAGGCTGTGTGATCCAGCTGGAGAGGAGGAGGAGGAGGAG 1968  
Db 1921 GTGAACACCCCGGCTGTGAGGCTGTGTGATCCAGCTGGAGAGGAGGAGGAGGAGGAG 1980  
Qy 1969 GCCGAGACCTTCTACGTGGAGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2028  
Db 1981 GCCGAGACCTTCTACGTGGAGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040

Qy 2029 TACGTGACCGACCGGGGCGGCGAGAGATCTGTAGCTCTGAACGAGACCAACCAAG 2088  
Db 2041 TACGTGACCGACCGGGGCGGCGAGAGATCTGTAGCTCTGAACGAGACCAACCAAG 2100  
Qy 2089 ACCGAGCTGACGGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2148  
Db 2101 ACCGAGCTGACGGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160  
Qy 2149 ACCGAGACCGAGTACGCGCTGGGCGATCATCAGGCGGCGGAGGAGGAGGAGGAGGAG 2208  
Db 2161 ACCGAGACCGAGTACGCGCTGGGCGATCATCAGGCGGCGGAGGAGGAGGAGGAGGAG 2220  
Qy 2209 CTGTGTAACGAGATCATCAGGAGCTGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2268  
Db 2221 CTGTGTAACGAGATCATCAGGAGCTGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280  
Qy 2269 CCGCGCCACAAAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2328  
Db 2281 CCGCGCCACAAAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340  
Qy 2329 CGCAAGGTGCTGTCTTCTGGAGCGGCGATCGATGGCGGCGATCGTGTATCTACGAGTACATGAG 2388  
Db 2341 CGCAAGGTGCTGTCTTCTGGAGCGGCGATCGATGGCGGCGATCGTGTATCTACGAGTACATGAG 2400  
Qy 2389 GACCTGTACGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2448  
Db 2401 GACCTGTACGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2460  
Qy 2449 GGTGAATTC 2457  
Db 2461 GGTGAATTC 2469

RESULT 4  
BD263705  
LOCUS  
DEFINITION BD263705 2300 bp DNA linear PAT 17-JUL-2003  
Improved expression of HIV polypeptides and production of virus-like particles.  
ACCESSION BD263705  
VERSION BD263705.1 GI:33073473  
KEYWORDS JP 2002533124-A/72.  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 2300)  
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: JP 2002533124-A 72 08-OCT-2002;  
CHIRON CORP  
COMMENT OS Artificial Sequence  
PN JP 2002533124-A/72  
PD 08-OCT-2002  
PR 30-DEC-1999 JP 2000591193  
PF 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471 PI  
SUSAN BARNETT, JAN ZUR MEDEDE, INDRESH SRIVASTAVA, YING LIAN, PI  
KARIN HARTOG,  
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC  
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC  
C12N5/10.  
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC  
Description of Artificial Sequence: FS(-). protomod.Rtopt.YMMH FH  
Key Location/Qualifiers  
FT source 1..2300  
FT Location/Qualifiers  
1..2300 /organism='Artificial Sequence'.  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'

FEATURES  
source  
ORIGIN

Query Match	83.0%;	Score 2040;	DB 6;	Length 2300;
Best Local Similarity	93.6%;	Pred. No. 4.6e-204;		
Matches 2153;	Conservative	0;	Mismatches 135;	Indels 12; Gaps 2;
170	GGGCAAGAGGGCCAC	CAGATGAAGGACTC	CACCGAGCGCCAGGCAACTTCTTCGCG	229
1	GGCGCGGAGAGCA	CCAAATGAAGATTGCAT	GTGAGACAGCGCTAATTTCTTCGCG	60
230	AGGACCTGGCTTCCC	CAGGGCAAGGC	CCGCGAGTTCCCCAGCGAGCAACCGGCGCA	289
61	AGGACCTGGCTTCTG	CAGGGCAAGGC	CCGCGAGTTCCAGCGAGCAGACCGGCGCA	120
290	ACAGCCCCACCA	CGCGCAGCTG	CAGGTGCGCGGG-----ACAACCCCGCAGCGAGG	343
121	ACAGCCCCACCG	CGCGCAGCTG	CAGGTGCGCGGGGAGAA	CAA
344	CGCGCGCGAGCG	CGCAGGGCACCTGTG	-----AACTTCCCCCAGATCACCTGTGGCAGC	397
181	CGCGCGCGAGCG	CGCAGGGCACCTGTG	AGCTTCAACTTCCCCCAGATCACCTGTGGCAGC	240
398	GGCCCCCTGGTGAG	CATCAAGGTGGCGCGC	AGATCAAGAGAGCCCTCTGTGACACCGGCG	457
241	GGCCCCCTGGTGAG	CATCAAGGTGGCGCGC	AGATCAAGAGAGCCCTCTGTGACACCGGCG	300
458	CGAGCGACACCTGT	CTGAGAGAGATGAGCT	CGCCGCGCAAGTGGAAGCCCAAGATGATCG	517
301	CGAGCGACACCTGT	CTGAGAGAGATGAGCT	CGCCGCGCAAGTGGAAGCCCAAGATGATCG	360
518	GGCGCATCGCGGCT	TTCATCAAGGTGGCGCGC	AGTAGCACAGATCCTTGATCGAGATCTGGG	577
361	GGCGCATCGCGGCT	TTCATCAAGGTGGCGCGC	AGTAGCACAGATCCTTGATCGAGATCTGGG	420
578	GCAAGAGCCATCG	CACCGTGTGATCGCGCCCA	CCCCCGTGAACATCATCGGCGCGCA	637
421	GCCAAGGCCATCG	CACCGTGTGATCGCGCCCA	CCCCCGTGAACATCATCGGCGCGCA	480
638	ACATGCTGACCCAG	CTGGGCTGCACTTCCCAT	CAGCCCCCATCGAGACCGGTGC	697
481	ACCTGCTGACCCAG	ATGGCTGCACTTCCCAT	CAGCCCCCATCGAGACCGGTGC	540
698	CGGTGAAGCTGA	AGCCCGGATGAGCGCCCA	AGAGGTGAAGCAGTGCCCTGACCGAGG	757
541	CGGTGAAGCTGA	AGCCCGGATGAGCGCCCA	AGAGGTGAAGCAGTGCCCTGACCGAGG	600
758	AGAGATCAAGSCCT	GACCGCATCTCGAGGAGAT	GGAAGAGGGGCAAGATCACCA	817
601	AGAGATCAAGSCCT	GATCTGTGAGATCTGCA	CCGAGATGGAAGAGGGCAAGATCAGCA	660
818	AGATCGGCCCGAG	AACCCCTACAAACCCCGCT	GTGTGCGCATCAAGAAAGAGACAGCA	877
661	AGATCGGCCCGAG	AACCCCTACAAACCCCGCT	GTGTGCGCATCAAGAAAGAGACAGCA	720
878	CCAAGTGGCGA	AGCTGTGAGCTTCCGCGAGCT	GAAACAGCGCACCCAGGACTTCTGGG	937
721	CCAAGTGGCGA	AGCTGTGAGCTTCCGCGAGCT	GAAACAGCGCACCCAGGACTTCTGGG	780
938	AGGTGAGCTGGG	CATCCCCACCCCGCGGCT	CGTGAAGAAAGAGCGGTGACCGGTGC	997
781	AGGTGAGCTGGG	CATCCCCACCCCGCGGCT	CGTGAAGAAAGAGCGGTGACCGGTGC	840
998	TGGAAGTGGGGA	CGCCTACTTACAGCTGTGCCCT	GTGAGAGGACTTTCGGCAGTACACCG	1057
841	TGGAAGTGGGGA	CGCCTACTTACAGCTGTGCCCT	GTGAGAGGACTTTCGGCAGTACACCG	900
1058	CTTCAACATCCC	CAGCATCAACACAGAGAC	CCCCCGCATCCGCTACAGGTACAACTGTC	1117
901	CTTCAACATCCC	CAGCATCAACACAGAGAG	CCCCCGCATCCGCTACAGGTACAACTGTC	960
1118	TGCCCCAGGGCT	GTGAAGGGCAGCCCCAGCAT	TTTCCAGAGCAGCATGACCAAGATCTCGG	1177
961	TGCCCCAGGGCT	GTGAAGGGCAGCCCCAGCAT	TTTCCAGAGCAGCATGACCAAGATCTCGG	1020

1178	Qy	AGCCCTTCGCGCCCGGCAACCCCGAGATCGTGATCTACAGAGGCCCCCTGTGACGTGGGCA	1231	Qy	AGCCCTTCGCGCCCGGCAACCCCGAGATCGTGATCTACAGAGGCCCCCTGTGACGTGGGCA
1021	Db	AGCCCTTCGCAAGCAGAAACCCCGACATCGTGATCTACAGAGGCCCCCTGTGACGTGGGCA	1080	Db	AGCCCTTCGCAAGCAGAAACCCCGACATCGTGATCTACAGAGGCCCCCTGTGACGTGGGCA
1238	Qy	GCGACCTGGAGATCGCCGACGACCGCGCCAAAGATCGAGGAGCTGCGCAACGACCTGCTGC	1297	Qy	
1081	Db	GCACCTGGAGATCGGCAGCAACCGCACCAAGATCGAGGAGCTGCGCCAGCATCTGCTGC	1140	Db	
1298	Qy	GCTGGGCTTTCCACACCCCGCAACAAAGCACACAGAGGAGGCCCTTCTGCCCATCG	1357	Qy	
1141	Db	GCTGGGCTTCAACACCCCGCAGAGNAGCACAGAGGAGGCCCTTCTGCCCATCG	1200	Db	
1358	Qy	AGTGCACCCCGACAGATGGAACCGTGAGCCCATCGAGCTGCCGAGAGAGGAGCTGGA	1417	Qy	
1201	Db	AGTGCACCCCGACAAGTGAACCGTGAGCCCATCATGTGTGCCCGAGAGGACAGCTGGA	1260	Db	
1418	Qy	CCGTGAACGACATCCAGAGCTGGTGGCAAGCTGAACCTGGGCGAGCCAGATCTACCCCG	1477	Qy	
1261	Db	CCGTGAACGACATCCAGAGCTGGTGGCAAGCTGAACCTGGGCGAGCCAGATCTACCGCG	1320	Db	
1478	Qy	GCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCGCAAGAGCCCTGACCGACATCG	1537	Qy	
1321	Db	GCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCACCAAGGCCCTGACCGAGTGA	1380	Db	
1538	Qy	TGCCCTTGACCGGAGGCGCGAGCTGAGCTGCGCGAGAAACCGCGAGATCCTTGGCGGAGC	1597	Qy	
1381	Db	TCCCTCTGACCGAGGAGCGCGAGCTGAGCTGCGCGGACCAAGGCCCTGACCGAGGAGC	1440	Db	
1598	Qy	CGTGCACGGCGTGTTACGACCCCGACNAGAGCACTGTTGGCCGAGATCCAGNAGCAGG	1657	Qy	
1441	Db	CCGTGCACGAGGTGTATTACGACCCCGACNAGAGCACTGTTGGCCGAGATCCNAGAGCAGG	1500	Db	
1658	Qy	GCCAGCACAGTGTGACCTTACCAGATCTACCAGAGGCCCTTCAAGAACCTTCAAGACCGGCA	1717	Qy	
1501	Db	GCCAGGCCAGTGTGACCTTACCAGATCTACAGAGGCCCTTCAAGAACCTTGAAGACCGGCA	1560	Db	
1718	Qy	AGTACGCAAGATGCGCACCGCCACCAACCGACGTGAAGCAGCTGACCGAGGCGGTGC	1777	Qy	
1561	Db	AGTAGCCCGCATGCGCGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCGGTGC	1620	Db	
1778	Qy	AGAAGATGCCATGGAGAGCATCGTGATCTGGGGCGAGACCCCAAGTTTCGGCTGCCCA	1837	Qy	
1621	Db	AGAAGTGAACACCGAGAGCATCGTGATCTGGGGCGAAGATCCCAAGTTTCAAGCTGCCCA	1680	Db	
1838	Qy	TCAGAGGAGAGACTGGGAGACCTGGTGGACGACTACTGGCAGGCCACCTGGATCCCCG	1897	Qy	
1681	Db	TCCAGAGGAGACCTGGGAGGCTTGTGTGATGTGAGTACTGCGAGGCCACCTGGATCCCCG	1740	Db	
1898	Qy	AGTGGGAGTTTCGTGAACACACCCCCCTCGTGAAGCTGTGGTACCACTGGAGAGGAGC	1957	Qy	
1741	Db	AGTGGGAGTTTCGTGAACACACCCCCCTCGTGAAGCTGTGGTACCAGCTGGAGAGGAGC	1800	Db	
1958	Qy	CCATCATCGCGCGGAGACCTTCTACGTGAACCGCGCGCCCAACCGCGAGACCAAGATCG	2017	Qy	
1801	Db	CCATCTGTGGCGCGAGACCTTCTACGTGAACCGCGCGCGCTCAACCGCGAGACCAAGCTGG	1860	Db	
2018	Qy	GCAAGGCCGCTACGTGACCGACCGGGCGCGCAGAGATCTGTGAGCTGACCGAGACCA	2077	Qy	
1861	Db	GCAAGGCCGCTACGTGACCGACCGGGCGCGCAGAGGTTGGTGAAGCATCGCCGACACCA	1920	Db	
2078	Qy	CCAACAGAAAGACCGAGCTCGAGGCCATCCAGCTTGGCCCTGCAAGGACAGCGGAGCGAGG	2137	Qy	
1921	Db	CCAACAGAAAGACCGAGCTCGAGGCCATCCACTTGGCCCTGCAAGGACAGCGGCTGGAGG	1980	Db	
2138	Qy	TGAACATCGTGAACGACAGCAGGTACGCCCTTGGGCATCATCCAGGCCCGCCGACAAAGA	2197	Qy	
1981	Db	TGAACATCGTGAACGACAGCAGGTACGCCCTTGGGCATCATCCAGGCCCGCCGACAAAGA	2040	Db	
2198	Qy	GCGNAGCGAGCTGGTGAAACAGATCATCAGCAGCTGATCAAGAGGAGAGGTTGTACC	2257	Qy	
2041	Db	GCGNAGCGAGCTGGTGAGCCAGATCATCAGCAGCTGATCAAGAGGAGAGGTTGTACC	2100	Db	
2258	Qy	TGAGCTGGGTGCCCGCCCAACAAGGGCATCGCGCGGCAACGAGCAGATCGACNAGCTGTGA	2317	Qy	

```
Db 2101 TGGCTGGTGCCTCCGCCACAAAGGCGCATCGCGGCAACGAGCAGTGACAAAGCTGTGTGA 2160
Qy 2318 GCAAGGGCATCGCAAGGTGCTGTTCTTGGACGGCATCGATGGCGGATCGTGTATCTACC 2377
Db 2161 GCGCGGATCGCGAAGGTGCTGTTCTTGAACGGCATCGATGGCGGATCGTGTATCTACC 2220
Qy 2378 AGTACATGACGACCTGTACGTGGGACGCGCGCCCTAGGATCGATTTAAAGCTTCCCG 2437
Db 2221 AGTACATGACGACCTGTACGTGGGACGCGCGCCCTAGGATCGATTTAAAGCTTCCCG 2280
Qy 2438 GGGCTAGCACCGGTGAATTC 2457
Db 2281 GGGCTAGCACCGGTGAATTC 2300

RESULT 5
CQ870575
LOCUS CQ870575 2300 bp DNA linear PAT 13-SEP-2004
DEFINITION Sequence 83 from Patent EP1433851.
ACCESSION CQ870575
VERSION CQ870575.1 GI:52000091
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL Patent: EP 1433851-A 83 30-JUN-2004;
CHIRON CORPORATION (US)
FEATURES
source
location/Qualifiers
1..2300
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description of Artificial Sequence:
FS(-).protmod.RTopt.YMMW"

ORIGIN

Query Match 83.0%; Score 2040; DB 6; Length 2300;
Best Local Similarity 93.6%; Pred. No. 4.6e-204;
Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

Qy 170 GCGCAAGGAGGCGCACGAGATGAGACTGCACCGAGCGCGCAGCCCACTTCTTCGCG 229
Db 1 GCGCGCGGAGGAGACCAAAATGAAGATTGCACTGAGAGACAGGCTAAATTTCTTCGCG 60
Qy 230 AGGACCTGGCTTCCCGCAGGCAAGGCGCGGAGTTCCCGAGGAGCAGAACCGCGCCA 289
Db 61 AGGACCTGGCTTCTGCAAGGCAAGGCGCGGAGTTGAGGAGGAGCAGACCGCGCCA 120
Qy 290 ACAGCCCAACAGCGCGAGCTGAGGTGCGCGCG-----ACAACCCCGCAGCGAGG 343
Db 121 ACAGCCCAACCGCGCGAGCTGAGGTGCGCGCGGAGAAACAAGCGCTGAGCGAGG 180
Qy 344 CCGCGCGGAGCGCGAGGCAACCTG-----ACCTTCCCGCAGATCACCTGTGGCAGC 397
Db 181 CCGCGCGGAGCGCGAGGCAACCTGAGCTTCAACTTCCCGCAGATCACCTGTGGCAGC 240
Qy 398 GCCCTCTGTGAGCATCAAGTGGCGGCGCAGATCAAGGAGGCGCTCTGTGACACCGCGC 457
Db 241 GCCCTCTGTGACCATCAGGATCGCGCGCGAGCTCAAGGAGGCGCTCTGTGACACCGCGC 300
Qy 458 CCGACGACACCGTCTGTGAGGAGATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGACGACACCGTCTGTGAGGAGATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCG 360
Qy 518 CCGGATCGCGGCTTCAATCAAGTGGCGGAGTACGACCGAGATCCTGATCGAGATCTGCG 577
Db 361 CCGGATCGCGGCTTCAATCAAGTGGCGGAGTACGACCGAGATCCCGGTGAGATCTGCG 420
```

```
Qy 578 GCAAGAAAGGCGCATCGGCAACCGTGTGATCGGCCCCACCCCGTGAACATCATCGGCCGCA 637
Db 421 GCCACAAAGGCATCGGCACCGTGTGTGGGCCCCACCCCGTGAACATCATCGGCCGCA 480
Qy 638 ACATGTGACCCAGCTGGGTGCAACCTTGAATTTCCCATCAGCCCCCATCGAGACGCTGC 697
Db 481 ACCTGTGACCCAGATCGGCTGCAACCTTGAATTTCCCATCAGCCCCCATCGAGACGCTGC 540
Qy 698 CCGTGAAGCTGAAGCGCGGATGAGCGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 757
Db 541 CCGTGAAGCTGAAGCGCGGATGAGCGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 600
Qy 758 AGAAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGATGGAAGAGGCGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCCCTTGTTGAGATCTGCACCGAGATGGAGAGGCGCAAGATCACCA 660
Qy 818 AGATCGGCGCGAGAAACCCCTTAACACCCCGCTGTGTCGCCATCAAGAAAGAGACAGCA 877
Db 661 AGATCGGCGCGAGAAACCCCTTAACACCCCGCTGTGTCGCCATCAAGAAAGAGACAGCA 720
Qy 878 CCAAGTGGCGCAGCTGTTGAGCTTCCGCGAGCTGAACAGCGCACCCAGACCTTCTGGG 937
Db 721 CCAAGTGGCGCAGCTGTTGAGCTTCCGCGAGCTGAACAGCGCACCCAGACCTTCTGGG 780
Qy 938 AGGTGACAGCTGGGCGATCCCCCACCCCGCGCTGAAGAAAGAGCGCTGACCGTGC 997
Db 781 AGGTGACAGCTGGGCGATCCCCCACCCCGCGCTGAAGAAAGAGCGCTGACCGTGC 840
Qy 998 TGAAGTGGGCGACGCTTACTTACAGTGGCCCCCTGGAAGAGGACTTCGCAAGTACACG 1057
Db 841 TGAAGTGGGCGACGCTTACTTACAGTGGCCCCCTGGAAGAGGACTTCGCAAGTACACG 900
Qy 1058 CCTTACCATCTCCCGAGATCAACAGAGACCCCGCGCTCCGCTACAGTACAGCGTGC 1117
Db 901 CCTTACCATCTCCCGAGATCAACAGAGACCCCGCGCTCCGCTACAGTACAGCGTGC 960
Qy 1118 TGCCCCAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1177
Db 961 TGCCCCAGGCTGGAAGGCGAGCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1020
Qy 1178 AGCCCTTCCGCGCGCCCAACCCCGAGATCTGTACTACAGGCCCCCTGTGACGTGGGCA 1237
Db 1021 AGCCCTTCCGCAAGACGAAACCCCGCATCTGTACTACAGGCCCCCTGTGACGTGGGCA 1080
Qy 1238 GCGACCTGGAGATCGGCGACGACCGCGCAAGATCGAGGAGCTGGCAAGCAGCTGCTGC 1297
Db 1081 GCGACCTGGAGATCGGCGACGACCGCGCAAGATCGAGGAGCTGGCGCAGCACCTGCTGC 1140
Qy 1298 GCTGGGCTTTCACCAACCCCGCAAGAAAGCAACCAAGAGGAGCCCCCTTCTGCGCCATCG 1357
Db 1141 GCTGGGCTTTCACCAACCCCGCAAGAAAGCAACCAAGAGGAGCCCCCTTCTGCGCCATCG 1200
Qy 1358 AGTGCACCCCGCAAGTGAACCGTGCAGCCATTCGAGCTGCCGAGAGGAGAGCTGGA 1417
Db 1201 AGTGCACCCCGCAAGTGAACCGTGCAGCCATTCATGCTGCGCGAGAGAGCAGCTGGA 1260
Qy 1418 CCGTGAACGACATCCAGAACTGTTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCG 1477
Db 1261 CCGTGAACGACATCCAGAACTGTTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCG 1320
Qy 1478 GCATCAAGGTGCGCAGCTGTCAAGCTGTGCGCGCGCCCAAGGCCCTGACCCAGCATCG 1537
Db 1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGGCAACCAAGGCCCTGACCCAGGTGA 1380
Qy 1538 TGCCCTTGACCGGAGGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCCTGCGCGAGC 1597
Db 1381 TCCCCTTGACCGGAGGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCCTGAGGAGC 1440
Qy 1598 CCGTGAACGCGGCTGTACTACGACCCCGAGAGGACCTGTTGGCGCGAGATCCAGAGCAGG 1657
Db 1441 CCGTGAACGAGGTGTACTACGACCCCGAGAGGACCTGTTGGCGCGAGATCCAGAGCAGG 1500
```

Qy	1658	GCCACGACCAAGTGGACCTACCAAGATCTACCAAGAGCCCTTCAAGAACCTTGAAGACCGCA	1717
Db	1501		1560
Qy	1718	AGTACGCCAAGATGGCCACCGCCCAACACACGACGTGAAGCAGCTGACCGAGCGCGTGC	1777
Db	1561		1620
Qy	1778	AGAAAGTCGCCATGAGAGACATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCA	1837
Db	1621		1680
Qy	1838	TCACAAGGAGACCTGGGAGACCTCGTGACCGACTACTTGGCAGGCCCACTTGGATCCCG	1897
Db	1681		1740
Qy	1898	AGTGGAGTTCGTGAACACCCCCCTCGTGAAGCTGTGTGTAACAGCTGGAGAGGAGC	1957
Db	1741		1800
Qy	1958	CCATCATGGCGCCGAGACCTTCTAGCTGGACGGCGCCGCAACCGCGAGACCAAGATCG	2017
Db	1801		1860
Qy	2018	GCAAGGCGGCTACGTGACCGACCGGGCCGGCAGAAGATCGTGAGCCTGACCCGAGACCA	2077
Db	1861		1920
Qy	2078	CCAACCAAGAAGCCGAGCTGCAAGGCCATCCAGCTGGCCCTGCAGGACACGCGCAGG	2137
Db	1921		1980
Qy	2138	TGAACATGTTGACACGACGACCTTCTAGCTGGACGGCGCTGGGCAATATCAGGCCAGCCCGACAGA	2197
Db	1981		2040
Qy	2198	GGGAGCGAGCTGTGTGAACCAAGATCATCGACAGCTGATCAAGAGAGAGGTGTACC	2257
Db	2041		2100
Qy	2258	TGAGCTGGGTGCCCCCAAGGGCATCGGGCGCAACGAGCAGATGCAACAGCTGGTGA	2317
Db	2101		2160
Qy	2318	GCAAGGGCATCCGCAAGTGTCTTCTGGAGCGCATCGATGGCGGCATCGTGAATCTACC	2377
Db	2161		2220
Qy	2378	AGTATATGACGACCTGTACGTGGGCGAGCGGGCCCTAGGATCGATTAAAGCTTCCCG	2437
Db	2221		2280
Qy	2438	GGGCTAGCACCGGTGAATTC 2457	
Db	2281	GGGCTAGCACCGGTGAATTC 2300	

RESULT 6  
AR373388  
LOCUS AR373388 2300 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 83 from patent US 6602705.  
ACCESSION AR373388  
VERSION AR373388.1 GI:40075491  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2300)  
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.  
TITLE Expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: US 6602705-A 83 05-AUG-2003;  
FEATURES Location/Qualifiers

		source	1. .2300 /organism="unknown" /mol_type="genomic DNA"
		ORIGIN	
		Query Match	83.0%; Score 2040; DB 6; Length 2300;
		Best Local Similarity	93.6%; Pred. No. 4.6e-204;
		Matches 2153; Conservative	0; Mismatches 135; Indels 12; Gaps 2;
Qy	170	GCGCAAGGAGGGCCACGATGAAGGACTGCAACGAGCGCCAGGCCAACTTCTTCGGG	229
Db	1		60
Qy	230	AGGACCTGGCTTCCCCCAGGCGAGGCCCGGAGTTCCCGAGTTCGCCAGCAGAACCCGCGCA	289
Db	61		120
Qy	290	ACAGCCCCACAGCGCGAGCTGCAAGTTCGCGCGG-----ACAACCCCGCAGCGAGG	343
Db	121		180
Qy	344	CGGCGCCGAGGCGCAGGCGACCCCTG-----AAGTCCCCAGATACCCCTGTGGGAGC	397
Db	181		240
Qy	398	GCCCTCTGGTCAAGTCAAGGTGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG	457
Db	241		300
Qy	458	CCGACGACACCGCTGTGGAGGAGATGAGCTGCCCGCAAGTGGAGGCCCAAGATGATCG	517
Db	301		360
Qy	518	GCGGATCGGCGGCTTATCAAGGTGGCGCCAGTAGCACAGATCTGATTCGAGATCTGGG	577
Db	361		420
Qy	578	GCAAGAGGCATCGGCACCGTGTGATCGCGCCACCCCGTGAACATCATCGGCCCA	637
Db	421		480
Qy	638	ACATGCTGACCCAGCTGGGCTGCACCTTGAATTTCCCATCAGCCCCATCGAGACCGTGC	697
Db	481		540
Qy	698	CCGTGAAGCTGAAGCGCGGATGGAACGCGCCCAAGGTCAAGGTGGGCGCTTGAACATCAT	757
Db	541		600
Qy	758	AGAGATCAAGCGCCTGACCGCCATCTCGAGGAGATGGAGAGGAGGCGCAAGATCACCA	817
Db	601		660
Qy	818	AGATCGGCCCGAGAACCCCTACAAACCCCGTGTTCGCCATCAAGAGAGGACAGCA	877
Db	661		720
Qy	878	CAAAGTGGCGAAGCTGTGTGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG	937
Db	721		780
Qy	938	AGGTGACGTGGGCATCCCGCCCGCGCTTGAAGAAGAGAGCGTGAACCGTGC	997
Db	781		840
Qy	998	TGGAGTGGGCGAGCCCTACTTTCAGCGTGCCTTCGACGAGACTTCCGCAAGTACACCG	1057
Db	841		900
Qy	1058	CCTTACCATCCGAGGATCAACAACGAGACCCCGGCTCCGCTACCGATACAGTACAGTGC	1117
Db	901		960

```
QY 1118 TGCCCGAGGGCTGAAGGGCGAGCCCGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1177
Db 961 TGCCCGAGGGCTGAAGGGCGAGCCCGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1020
QY 1178 AGCCCTTCCGCGCCGCGCAACCCCGAGATCGTGATCTACCAAGCCCGCCCTGTGACGTGGGCA 1237
Db 1021 AGCCCTTCCGCGAGCAGNACCCCGACATCGTGATCTACCAAGCCCGCCCTGTGACGTGGGCA 1080
QY 1238 GCGACCTGGAGATCGGCGAGCACCGCGCCAGATCGAGGAGCTGCGAAGCACCTGCTGTC 1297
Db 1081 GCGACCTGGAGATCGGCGAGCACCGCGCCAGATCGAGGAGCTGCGCGAGCACCTGCTGTC 1140
QY 1298 GCTGGGGCTTACACACCCCGAGCAAGAAGACNACAGAGGAGCCCCCTTCTGCGCCATCG 1357
Db 1141 GCTGGGGCTTACACACCCCGAGCAAGAAGACNACAGAGGAGCCCCCTTCTGCGCCATCG 1200
QY 1358 AGCTGACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGA 1417
Db 1201 AGCTGACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGA 1260
QY 1418 CCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1477
Db 1261 CCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCG 1320
QY 1478 GCATCAAGGTGGCCAGCTGTGCAAGCTGTGCGCGCGCCAGGCGCTGACGACATCG 1537
Db 1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCCAGGCGCTGACGAGGTGA 1380
QY 1538 TGCCCTGACCGAGGAGCGAGCTGGAGCTGGCGGAGAACCGGAGATCTGCGCGAGC 1597
Db 1381 TGCCCTGACCGAGGAGCGAGCTGGAGCTGGCGGAGAACCGGAGATCTGCGCGAGC 1440
QY 1598 CCGTGAACGAGTGTACTACGACCCCGAGCAAGCACTGTGGGCCAGATCCAGAAGCAGG 1657
Db 1441 CCGTGAACGAGTGTACTACGACCCCGAGCAAGCACTGTGGGCCAGATCCAGAAGCAGG 1500
QY 1658 GCGACGACGAGTGACCTACAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCA 1717
Db 1501 GCGAGGCGAGTGAGACCTACAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCA 1560
QY 1718 AGTACGCCAAGATCGCGACCGCCACACACGACGAGTGAAGCAGCTGACCGAGCGCTGC 1777
Db 1561 AGTACGCCGAGATCGCGGCGCCACACACGACGAGTGAAGCAGCTGACCGAGCGCTGC 1620
QY 1778 AGAAGATCGCATGGAGAGCATCTGTACTGGGCAAGACCCCAAGTTCGCGCTGCCCA 1837
Db 1621 AGAAGTGAGCACCGAGAGCATCTGTACTGGGCAAGATCCCAAGTTCAGCTGCCCA 1680
QY 1838 TCCAGAGGAGACCTGGGAGACCTGTGGACCGGACCTACTGGCAGGCGACCTGGATCCCG 1897
Db 1681 TCCAGAGGAGACCTGGGAGGCGCTGTGGATGGAGTACTGGCAGGCGACCTGGATCCCG 1740
QY 1898 AGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACCACTGGAGAGGAGC 1957
Db 1741 AGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACCACTGGAGAGGAGC 1800
QY 1958 CCATCATCGCGCGAGACCTTCTACTGTGAACGCGCGCCGCAACCGGAGACCAAGATCG 2017
Db 1801 CCATCGTGGGCGCGAGACCTTCTACTGTGAACGCGCGCCGCAACCGGAGACCAAGCTGG 1860
QY 2018 GCAAGGCGGCTACGTGAACCGGCGCGGCGGCAAGAGATCGTGAGCTGACCGGAGACCA 2077
Db 1861 GCAAGGCGGCTACGTGAACCGGCGCGGCGGCAAGAGTGGTGAAGCATCGCGGAGACCA 1920
QY 2078 CCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTGGGCGCTGCAGGACGCGGCGAGG 2137
Db 1921 CCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTGGGCGCTGCAGGACGCGGCGTGGAG 1980
QY 2138 TGAACATCGTGACCGGACGAGCAGTACGCGCTTGGGCGATCATCCAGGCGCGGCGGCAAGA 2197
Db 1981 TGAACATCGTGACCGGACGAGCAGTACGCGCTTGGGCGATCATCCAGGCGCGGCGGCAAGA 2040
QY 2198 GCGAGAGCGAGCTGGTGAACCGAGATCATCGAGCAGCTGATCAAGAGAGGAGGTGTACC 2257
```

```
Db 2041 GCGAGAGCGAGCTGGTGGAGCCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACC 2100
QY 2258 TGAGCTGGGTGCCCGCCCAACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGA 2317
Db 2101 TGCGCTGGGTGCCCGCCCAACAAGGGCATCGGCGGCAACGAGCAGGTGGACAACTGGTGA 2160
QY 2318 GCAAGGCGATCCGCAAGTGTCTCTGTCGACGCGCATCGATGCGGCGCATCTGTATACC 2377
Db 2161 GCGCGGCGATCCGCAAGTGTCTCTCTGAAACGCGCATCGATGCGGCGCATCTGTATACC 2220
QY 2378 AGTACATGAGGAGCCTGTAGCTGGGCGAGCGCGCGCTAGGATCGATTAAGCTTCCCG 2437
Db 2221 AGTACATGAGGAGCCTGTAGCTGGGCGAGCGCGCGCTAGGATCGATTAAGCTTCCCG 2280
QY 2438 GGGCTAGCACCGGTGAATTC 2457
Db 2281 GGGCTAGCACCGGTGAATTC 2300

RESULT 7
BD263704
LOCUS
DEFINITION BD263704 2306 bp DNA linear PAT 17-JUL-2003
Improved expression of HIV polypeptides and production of
virus-like particles.
ACCESSION BD263704
VERSION BD263704.1 GI:33073472
KEYWORDS JP 2002533124-A/71.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 2306)
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL Patent: JP 2002533124-A 71 08-OCT-2002;
CHIRON CORP
COMMENT
OS Artificial Sequence
PN JP 2002533124-A/71
PD 08-OCT-2002
PF 30-DEC-1999 JP 2000591193
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI
SUSAN BARNETT,JAN ZUR MEDEDE,INDRESH SRIVASTAVA,YING LIAN, PI
KARIN HARTOG,
PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
C12N5/10.
PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH Key
Location/Qualifiers
FT source 1..2306
FT /organism='Artificial Sequence'.
FEATURES
source Location/Qualifiers
1..2306
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 82.2%; Score 2019.2; DB 6; Length 2306;
Best Local Similarity 93.2%; Pred. No. 6.8e-202;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;
QY 170 GCGGCAAGGAGGCGCCACCATGAAGCACTGCACCGAGCGCCAGCCCACTTCTTCGCG 229
Db 1 GCGGCGCGGAGGAGCACCAATGAAGATTGCACTGAGAGACAGGCTTAATTTCTTCGCG 60
QY 230 AGGACCTGGCTTCCCGCAGGCAAGCGCCCGAGTTCCCGAGGAGCAGCAACCGCGCA 289
Db 61 AGGACCTGGCTTCTTCGAGGCGAAGCGCCCGGAGTTTCAGGCGGAGCAGACCGCGCA 120
QY 290 ACAGCCCCACAGCGCGAGCTGCGGCGCG-----ACAACCCCCGAGCGAGG 343
```



[illegible]

Db	1201	GCTACGAGCTGCACCCCGCAACAGTGGACCGTGCAGCCCATCATGCTGCCCGAAGGACA	1261
Qy	1412	GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCCAGATCT	1471
Db	1261	GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCCAGATCT	1320
Qy	1472	ACCCCGGACATCAAGGTGCGCAGCTGTGCAAGCTGTGCGCGCGCCAAAGGCCCTGACCG	1531
Db	1321	ACGCGGGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGGCACCAAGGCCCTGACCG	1380
Qy	1532	ACATGTGCCCTGACCGAGGAGCGCAGCTGAGCTGGCGGAGAACCGGCGAGATCCTGC	1591
Db	1381	AGGTGATCCCCCTGACCGAGGAGCGCAGCTGAGCTGGCGCGAGAACCGGAGATCCTGA	1440
Qy	1592	GCAGGCCGTGCACGGCGTGCTACTACGACCCAGCAAGGACCTGGTGGCGGAGATCCAGA	1651
Db	1441	AGGAGCCGTGCACGAGGTGCTACTGACCCAGCAAGGACCTGGTGGCGGAGATCCAGA	1500
Qy	1652	AGCAGGCCACGACCAAGTGGACCTACCAGATCTACAGGAGCCCTTCAAGAACCTGAAGA	1711
Db	1501	AGCAGGCCACGAGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA	1560
Qy	1712	CCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACAGAGCTGAAGCAGCTGACCGAGG	1771
Db	1561	CCGGCAAGTACGCCCGCATCGCGGGCGCCACACCAACAGAGCTGAAGCAGCTGACCGAGG	1620
Qy	1772	CCGTGCAGAGATCGCCATCGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCC	1831
Db	1621	CCGTGCAGAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCAAGTTCAAGC	1680
Qy	1832	TGCCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGCATACTGCGAGGCCACCTGGA	1891
Db	1681	TGCCCATCCAGAAAGGAGACCTGGGAGCCCTGGTGGATGGAGTACTGCGAGGCCACCTGGA	1740
Qy	1892	TCCCGGAGTGGGAGTTCTGTGAACACCCCCCTCTGTGTGAAGCTGTGGTACCAAGTGGAGA	1951
Db	1741	TCCCGGAGTGGGAGTTCTGTGAACACCCCCCTCTGTGTGAAGCTGTGGTACCAAGTGGAGA	1800
Qy	1952	AGAGGCCCATCATCGGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCGAGACCA	2011
Db	1801	AGAGGCCCATCTGTGGGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCGAGACCA	1860
Qy	2012	AGATCGGCAAGGCGCGCTACGTACCGACGACGGGGCCGGCAGAAAGATCGTGAGCCTGACCG	2071
Db	1861	AGCTGGGCAAGGCGCGCTACGTACCGACGACGGGGCCGGCAGAAAGTGGTGAGCATCGCCG	1920
Qy	2072	AGACCAACCAACGAGAGACCGAGCTGAGGCCATCCAGTGGCCCTGCGAGGACAGCGGCA	2131
Db	1921	ACACCAACCAACGAGAGACCGAGCTGCGAGGCCATCCACCTGGCCCTGCGAGGACAGCGGCC	1980
Qy	2132	GCAGGTGGAACATCGTGCACGACGACGATACGCCCTTGGGCATCATCCAGGCCACGCGCG	2191
Db	1981	TGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTTGGGCATCATCCAGGCCACGCGCG	2040
Qy	2192	ACAAGCGAGAGCGAGCTGGTGAACACGATCATCGAGCAGCTGATCAAGAAAGGAGAAGG	2251
Db	2041	ACAAGCGAGAGCGAGCTGGTGAAGCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAAGG	2100
Qy	2252	TGTACTGAGCTGGTGGCCGCCCAAGGGGCATCGCGGCAACGAGCAGATCGACAAAGC	2311
Db	2101	TGTACTTGGCTCGGTGTCGCCGCCCAAGGGGCATCGCGGCAACGAGCAGGTGGACAAAGC	2160
Qy	2312	TGTTGAGCAAGGCAATCCGCAAGGTGCTGTTCTTGGACGGCATCGATGGCGGATCGTGA	2371
Db	2161	TGTTGAGCGCCGGCATCCGCAAGGTGCTGTTCTTGAACGGCATCGATGGCGGATCGTGA	2220
Qy	2372	TCTACCAAGTACATGGACGACCTGTACTGGGCGAGCGCGGCCCTAGGATCGATTAAAGC	2431
Db	2221	TCTACCAAGTACATGGACGACCTGTACTGGGCGAGCGCGGCCCTAGGATCGATTAAAGC	2280
Qy	2432	TTCCCGGGGCTAGCACCGCGTGAATTC	2457
Db	2281	TTCCCGGGGCTAGCACCGCGTGAATTC	2306

RESULT 8  
CQ870574 2306 bp DNA linear PAT 13-SEP-2004  
LOCUS  
DEFINITION Sequence 82 from Patent EP1433851.  
ACCESSION CQ870574  
VERSION CQ870574.1 GI:52000090  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE  
AUTHORS Barnett, S., Zurnegede, J., Srivastava, I., Lian, Y., Hartog, K.,  
Liu, H., Greer, C., Selby, M. and Walker, C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: EP 1433851-A 82 30-JUN-2004;  
CHIRON CORPORATION (US)  
FEATURES  
source  
1. 2306  
Location/Qualifiers  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Description of Artificial Sequence:  
FS(-).protmod.RTopt.YM"

ORIGIN  
Query Match 82.2%; Score 2019.2; DB 6; Length 2306;  
Best Local Similarity 93.2%; Pred. No. 6.8e-202;  
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
QY 170 GCGCAAGAGGAGGCGCACAGATGAAGGACTGCACCGAGCGCGAGCCCAACTTCTTCGCG 229  
DB 1 GCGCGCGGAGGACACCAATGAAGATTGCATGAGAGCAGGCTAATTTCTTCGCG 60  
QY 230 AGGACTGGGCTTCCCGCAGGCAAGGCGCGAGTTCCCGAGTTCGCCAGCGAGCAGACCGGCCA 289  
DB 61 AGGACTGGGCTTCTCGAGGCAAGGCGCGAGTTTCAGCAGCGAGCAGACCGGCCA 120  
QY 290 ACAGCCCCACGCGGAGTGCAGGTGGCGGCGG-----ACAAACCGCGGAGCGAGG 343  
DB 121 ACAGCCCCACCGCGGAGTGCAGGTGGCGGCGGAGGAGAAACAGCGCTGAGCGAGG 180  
QY 344 CCGCGCGCGAGCGCGAGGCAACCGTGAGCTTCACTTCCCGCAGATCACCTGTGGCAGC 397  
DB 181 CCGCGCGCGAGCGCGAGGCAACCGTGAGCTTCACTTCCCGCAGATCACCTGTGGCAGC 240  
QY 398 GCGCGCGCGAGCGCGAGGCGCGAGGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCG 457  
DB 241 GCGCGCGCGAGCGCGAGGCGCGAGGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCG 300  
QY 458 CCGAGGACCGGCTGCTGAGGAGATGAGCTTCCCGCGAGGTGAAGCGCGAGATGATCG 517  
DB 301 CCGAGGACCGGCTGCTGAGGAGATGAGCTTCCCGCGAGGTGAAGCGCGAGATGATCG 360  
QY 518 GCGCGATCGCGGCTTTCATCAAGGTGGCGAGTACGACCGAGTGCAGATCTGCG 577  
DB 361 GCGGATCGCGGCTTTCATCAAGGTGGCGAGTACGACCGAGTGCAGATCTGCG 420  
QY 578 GCAAGAGGCGCATCGGCGACCGTGTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCA 637  
DB 421 GCCAAGGCGCATCGGCGACCGTGTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCA 480  
QY 638 ACATGCTGACCGAGTGGGTGCAACCTTGAATTTCCCGCATCAGCCCCCATCGAGACCGTGC 697  
DB 481 ACCTGCTGACCGAGTGGGTGCAACCTTGAATTTCCCGCATCAGCCCCCATCGAGACCGTGC 540  
QY 698 CCGTGAAGCTGAGCGCGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757  
DB 541 CCGTGAAGCTGAGCGCGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
QY 758 AGAAGATCAAGGCGCGCTGACCGCCATCTGCGAGGAGATGGAGAGGCGGCAAGATCACCA 817

601 AGAAGATCAAGGCGCGCTGCTGGAGATCTGCACCGAGATGGAAGAGGCGCAAGATCAGCA 660  
QY 818 AGATCGCGCGCGGAAACCCCTTACAAACCCCGGTTTCGCCATCAAGAGAGGAGGACAGCA 877  
DB 661 AGATCGCGCGCGGAAACCCCTTACAAACCCCGGTTTCGCCATCAAGAGAGGAGGACAGCA 720  
QY 878 CCAAGTGGCGCAAGCTGCTGGAGCTTCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGG 937  
DB 721 CCAAGTGGCGCAAGCTGCTGGAGCTTCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGG 780  
QY 938 AGTGTAGCTGGGATCTCCCGCAACCCCGCGGCTTGAAGAGAAAGAGCGTGCACCGTGC 997  
DB 781 AGTGTAGCTGGGATCTCCCGCAACCCCGCGGCTTGAAGAGAAAGAGCGTGCACCGTGC 840  
QY 998 TGGACGTGGGCGAGCGCTTACTTTCAGCGTGCCTTGGAGGAGCTTCCGCAAGTACACCG 1057  
DB 841 TGGACGTGGGCGAGCGCTTACTTTCAGCGTGCCTTGGAGGAGCTTCCGCAAGTACACCG 900  
QY 1058 CTTTACCATCCCGAGCATCAACAAACAGAGACCCCGCGCATCCGCTACCAAGTACCAACGTGC 1117  
DB 901 CTTTACCATCCCGAGCATCAACAAACAGAGACCCCGCGCATCCGCTACCAAGTACCAACGTGC 960  
QY 1118 TGCCCCAGGGTGAAGGGGAGCGCCAGCATCTTTCAGAGAGCATGATGACCAAGATCTCTGG 1177  
DB 961 TGCCCCAGGGTGAAGGGGAGCGCCCGCATCTTTCAGAGAGCATGATGACCAAGATCTCTGG 1020  
QY 1178 AGCCCTTCCGCGCGCGCAACCCCGAGATCTGTATACCAAGGCGCGCGCTGTACGTGGCA 1237  
DB 1021 AGCCCTTCCGCGCGCGCAACCCCGAGATCTGTATACCAAGGCGCGCGCTGTACGTGGCA 1080  
QY 1238 GCGACCTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTCGCAAGCACCTGTCTGC 1297  
DB 1081 GCGACCTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTCGCAAGCACCTGTCTGC 1140  
QY 1298 GCTGGGCTTTCACACCCCGCAAGAAAGCAACCAAGAGGAGCGCGCTTCTGTGCCAT-- 1355  
DB 1141 GCTGGGCTTTCACACCCCGCAAGAAAGCAACCAAGAGGAGCGCGCTTCTGTGGATGG 1200  
QY 1356 ----CGAGCTGCAACCGCGCAAGTGGACCGTGCAGCGCCATCGAGCTGCCGAGAGGAGA 1411  
DB 1201 GCTACGAGCTGCAACCGCGCAAGTGGACCGTGCAGCGCCATCATGCTGCCCGAGAGGACA 1260  
QY 1412 GCTGGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCGAGCAGATCT 1471  
DB 1261 GCTGGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCGAGCAGATCT 1320  
QY 1472 ACCCGGCGATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCG 1531  
DB 1321 ACGCGGCGATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCG 1380  
QY 1532 ACATCGTGGCTGACCGAGGAGCGAGCTGGAGCTGGCGCGAGAAACCGCGAGATCTCTGC 1591  
DB 1381 AGTGTATCCCGCTGACCGAGGAGCGAGCTGGAGCTGGCGCGAGAAACCGCGAGATCTCTGA 1440  
QY 1592 GCGAGCGCGTGCAGCGGCTGTACTACGACCGCGAGAGGAGCTGTGGCGCGAGATCCAGA 1651  
DB 1441 AGAGCGCGTGCAGCGAGTGTACTACGACCGCGAGAGGAGCTGTGGCGCGAGATCCAGA 1500  
QY 1652 AGCAGGCGCGACGACCGAGTGCACCTTACAGATCTACAGGAGCGCTTCAAGAACCTGAAGA 1711  
DB 1501 AGCAGGCGCGAGGCGCAGTGAGCTTACAGATCTACAGGAGCGCTTCAAGAACCTGAAGA 1560  
QY 1712 CCGCGAAGTACCGCAAGATGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1771  
DB 1561 CCGCGAAGTACCG 1620  
QY 1772 CCGTGAAGAGATCGCGCATGCGAGGAGCTGTGATCTGGGCGCAAGACCGCGCGCGCGCG 1831  
DB 1621 CCGTGAAGAGATGAGCAGCGAGGAGCTGTGATCTGGGCGCAAGATCTCCCAAGTTCAAGC 1680  
QY 1832 TGCCCATCCAGAGAGGAGACCTGGGAGACCTGTGGAGCCGACTACTTGGCAGCGCACCTTGA 1891

Db 1681 TGCCCATCCAGAAGGAGACCTGGAGGCGCTGGTGTGATGAGTACTGCGAGGCGCACTGGGA 1740  
Qy 1892 TCCCGAGTGGGAGTTCTGTGAACCCCGCCCTCGTGAAGCTGTGGTACAGCTGGAGA 1951  
Db 1741 TCCCGAGTGGGAGTTCTGTGAACCCCGCCCTCGTGAAGCTGTGGTACAGCTGGAGA 1800  
Qy 1952 AGGAGCCCATCATCGCGCGCCGAGACCTTCTAGCTGGAGCGGCCCGCCCAACCGCGAGACCA 2011  
Db 1801 AGGAGCCCATCGTGGGCGCGGAGACCTTCTAGCTGGAGCGGCCCGCCCAACCGCGAGACCA 1860  
Qy 2012 AGATCGGCAAGCGCGCTGTGTAACCGACCGGGCGCGGAGAGATCTGTAGCTTGACCG 2071  
Db 1861 AGCTGGGCAAGCGCGCTGTGTAACCGACCGGGCGCGGAGAGATCTGTAGCTTGACCG 1920  
Qy 2072 AGACCAACCAACAGAGACCGAGCTGCAAGGCCATCCAGCTGGCTGCGAGGACAGCGGCA 2131  
Db 1921 ACACCAACCAACAGAGACCGAGCTGCAAGGCCATCCAGCTGGCTGCGAGGACAGCGGCC 1980  
Qy 2132 GCGAGGTGAACATCGTGAACCGACAGCCAGTACGCCCTGGGCATCATCAGGCCCGAGCCCG 2191  
Db 1981 TGGAGGTGAACATCGTGAACCGACAGCCAGTACGCCCTGGGCATCATCAGGCCCGAGCCCG 2040  
Qy 2192 ACAAGAGCGAGAGCGAGCTGTGTGAACCGACATCATCGAGCAGCTGATCAAGAGGAGAAAG 2251  
Db 2041 ACAAGAGCGAGAGCGAGCTGTGTGAACCGACATCATCGAGCAGCTGATCAAGAGGAGAAAG 2100  
Qy 2252 TGTACTGTGCTGGTGGCGCGCCACAGGGGATCGGGGCGACAGCAGATCGACAGC 2311  
Db 2101 TGTACTGTGCTGGTGGCGCGCCACAGGGGATCGGGGCGACAGCAGATCGACAGC 2160  
Qy 2312 TGGTGAAGCAAGGGCATCGCAAGGTGCTGTTCTCGAGCGGCATCGATCGCGGCATCGTGA 2371  
Db 2161 TGGTGAAGCGCGGCATCGCAAGGTGCTGTTCTCGAGCGGCATCGATCGCGGCATCGTGA 2220  
Qy 2372 TCTACAGTACATGACGACCTGTGTAGTGGGAGCGCGCGCCCTAGGATCGATTAAAGC 2431  
Db 2221 TCTACAGTACATGACGACCTGTGTAGTGGGAGCGCGCGCCCTAGGATCGATTAAAGC 2280  
Qy 2432 TTCCCGGGCTAGCACCGGTGAATTC 2457  
Db 2281 TTCCCGGGCTAGCACCGGTGAATTC 2306

RESULT 9  
AR373387  
LOCUS AR373387 2306 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 82 from patent US 6602705.  
ACCESSION AR373387  
VERSION AR373387.1 GI:40075490  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 2306)  
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.  
TITLE Expression of HIV polypeptides and production of virus-like  
PARTICLES  
JOURNAL Patent: US 6602705-A 82 05-AUG-2003;  
FEATURES  
Location/Qualifiers  
source 1..2306  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 82.2%; Score 2019.2; DB 6; Length 2306;  
Best Local Similarity 93.2%; Pred. No. 6.8e-202;  
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
Qy 170 GCGCAAGGAGGCGCCACAGATGAAGGACTGCACCGAGCGCCAGCGCCAACTTCTTCGCG 229  
Db 1 GCGGCGGAGGAGCACCAAAATGAAGATTGCACTGAGAGACAGAGCTAAATTTCTTCGCG 60  
Qy 230 AGGACTTGGCTTCCCGGAGGCAAGGCGCGGAGTTCCCGAGCGAGCAACCGCGGCA 289

Db 61 AGGACTTGGCTTCTCTGCAAGGCGCGAGTTCAAGCAGGAGCAGACCCCGGCA 120  
Qy 290 ACAGCCCAACAGCGCGAGCTGCAGGTGCGGGG-----ACAAACCCCGCAGGAG 343  
Db 121 ACAGCCCAACCGCGCGAGCTGCAGGTGCGGGGCGGAGAACAAAGCTTGAGCGAGG 180  
Qy 344 CCGGCGCGAGCGCGCAGGCGCACCTG-----AACTTCCCAGATCACCTGTGGCAGC 397  
Db 181 CCGGCGCGAGCGCGCAGGCGCACCTGTAGCTTCACTTCCCAGATCACCTGTGGCAGC 240  
Qy 398 GCGGCTGTGAGCATCAAGGTGGCGCGCAGATCAAGGAGGCCCTGTGGACACCGGCG 457  
Db 241 GCGGCTGTGAGCATCAAGGTGGCGCGCAGCTCAAGGAGGCGCTGTGCACACCGGCG 300  
Qy 458 GCGAGCACCGTGTGAGGAGATGAGCTTGCAGCGCAAGTGGAGGCCCAAGATGATCG 517  
Db 301 CCGAGCACCGTGTGAGGAGATGAACTTGCAGCGCAAGTGGAGGCCCAAGATGATCG 360  
Qy 518 GCGGCTGTGAGCATCAAGGTGGCGCGCAGTACGACAGATCCTGATCGAGATCTGCG 577  
Db 361 GCGGATCGGGGCTTCAAGGTGGCGCAGTACGACAGATCCCCTGTGGAGATCTGCG 420  
Qy 578 GCAAGAGGCCATCGGCAACCGTGTGATCGGCGCCCAACCGCTGAACATCATCGCGCGCA 637  
Db 421 GCCAAGGCCATCGGCAACCGTGTGATCGGCGCCCAACCGCTGAACATCATCGCGCGCA 480  
Qy 638 ACATGCTGACCGAGCTGGGCTGCACCTTGAACCTTCCCATCAGGCCCATCGAGACGCTGC 697  
Db 481 ACCTGCTACCGAGATCGGCTGCACCTTGAACCTTCCCATCAGGCCCATCGAGACGCTGC 540  
Qy 698 CCGTGAAGTGAAGCGCGCATGGAAGCGCCCAAGGTGAAGCAGTGGGCCCTGACCGGAG 757  
Db 541 CCGTGAAGTGAAGCGCGGATGGAAGCGCCCAAGGTCAAGCAGTGGGCCCTGACCGGAG 600  
Qy 758 AGAAGATCAAGGCCCTGACCGCCATCTCGAGGAGATGGAGAGGAGGCGCAAGATCAACA 817  
Db 601 AGAAGATCAAGGCCCTGATGGAGATCTGCAACCGAGATGGAGAGGAGGCGCAAGATCAACA 660  
Qy 818 AGATCGGCGCGAGAACCCCTTACAACACCCCGGTGTTGCCATCAAGAGAGGACAGCA 877  
Db 661 AGATCGGCGCGAGAACCCCTTACAACACCCCGGTGTTGCCATCAAGAGAGGACAGCA 720  
Qy 878 CCAAGTGGCGCAAGCTGTGAGCTTCCGCGAGCTGAACAAGCGCACCCAGAGCTTCTGGG 937  
Db 721 CCAAGTGGCGCAAGCTGTGAGCTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780  
Qy 938 AGGTGAGCTGGGCAATCCCCCAACCGCGCGCTGAAGAAGAGAGCGTGAACGCTGC 997  
Db 781 AGGTGAGCTGGGCAATCCCCCAACCGCGCGCTGAAGAAGAGAGCGTGAACGCTGC 840  
Qy 998 TGGAGTGGGCGAGCGCTTACTTTCAGCGTGGCGCGCTGGAAGGAGTACACCG 1057  
Db 841 TGGAGTGGGCGAGCGCTTACTTTCAGCGTGGCGCGCTGGAAGGAGTACACCG 900  
Qy 1058 CTTTCACCATCCCGAGCATCAACACCGAGACCCCGGATCCGCTACAGTACAAAGCTGC 1117  
Db 901 CTTTCACCATCCCGAGCATCAACACCGAGACCCCGGATCCGCTACAGTACAAAGCTGC 960  
Qy 1118 TGCCCCAGGCTGGAAGGCGAGCGCCAGCATTTTCCAGAGCAGCATGACCAAGATCTCTGG 1177  
Db 961 TGCCCCAGGCTGGAAGGCGAGCGCCCGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1020  
Qy 1178 AGCCCTTCGCGCGCGCAACCCCGAGATCGTGATCTACAGGCCCGCTGTACGTTGGCA 1237  
Db 1021 AGCCCTTCGCAAGAGAGAACCCCGCATCTGATCTACAGGCCCGCTGTACGTTGGCA 1080  
Qy 1238 GCGAGCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGC 1297  
Db 1081 GCGAGCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGAGCTGCGCAGCACTGCTGC 1140  
Qy 1298 GCTGGGCTTCAACACCCCGGCAAGAGCGCACCAAGAGGAGCGCCCGCTTCTGCGCCAT-- 1355



QY	602	TGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATCTCTGACCCAGCTGGGCTGCA	661
DB	3560	TGTTGGGGCCCCACCCCGTGAACATCATCGCGCGCAACCTGCTGACCCAGATCGGCTGCA	3619
QY	662	CCCTGAATCTCCCATCAGCCCCATCAGACACGCTGCCCGTGAAGCTGAAGCCCGGATGG	721
DB	3620	CCCTGAATCTCCCATCAGCCCCATCAGACACGCTGCCCGTGAAGCTGAAGCCCGGATGG	3679
QY	722	ACGSCCCCAAGTGAAGCAGTGGCCCCCTGACCGAGGAGAAAGATCAAGGCCCTGCACGCCA	781
DB	3680	ACGSCCCCAAGTGAAGCAGTGGCCCCCTGACCGAGGAGAAAGATCAAGGCCCTGGTGAGA	3739
QY	782	TCTCGAGGAGATGGAGAAGAGGGCAAGATCACCAAGATCGGCCCGCGAAGCCCTTACA	841
DB	3740	TCTGCAACCGAGATGGAGAAGAGGGCAAGATCAGCAAGATCGSCCCCGGAAACCCCTACA	3799
QY	842	ACACCCCGTGTTCGCCATCAAGAAGAGCAGACCAAGTGCGCGCAAGCTGGTGACT	901
DB	3800	ACACCCCGTGTTCGCCATCAAGAAGAGCAGACCAAGTGCGCGCAAGCTGGTGACT	3859
QY	902	TCCCGGAGCTGAAACAAGCGCACCCAGACATCTCGGAGGTGCAAGCTGGGCGATCCCCACC	961
DB	3860	TCCCGGAGCTGAAACAAGCGCACCCAGACATCTCGGAGGTGCAAGCTGGGCGATCCCCACC	3919
QY	962	CCGCGGCGCTGAAGAAGAAGAGCGGTGACCGCTGTGGAAGTGCGGCGACGCCTACTTCA	1021
DB	3920	CCGCGGCGCTGAAGCAGAGAGAGCGGTGACCGCTGTGGAAGTGCGGCGACGCCTACTTCA	3979
QY	1022	CGTGTGCCCTTGAACGAGATCTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAACA	1081
DB	3980	CGTGTGCCCTTGAACGAGATCTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAACA	4039
QY	1082	ACGAGACCCCGGGATCGCTACAGTACAAAGTGTGCGCCAGGGCTGGAAGGGCAGCC	1141
DB	4040	ACGAGACCCCGGGATCGCTACAGTACAAAGTGTGCGCCAGGGCTGGAAGGGCAGCC	4099
QY	1142	CCAGCATCTTCACAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGGCCGCAACCCCG	1201
DB	4100	CCGCCATCTTCAGTGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGAGCAGAACCCCG	4159
QY	1202	AGATCGTGATCTACCA-----GGCCCCCTGTACGTGGGCGAGCACTGGAGATCGGCC	1255
DB	4160	ACATCGTGATCTACCAAGTACATGGACCACTGTACGTGGGCGAGCACTGGAGATCGGCC	4219
QY	1256	AGCACCGGCCAAGATCGAGAGCTGCGAAGCAGCACTGCTGCGTGGGCTTCAACACCC	1315
DB	4220	AGCACCGGCCAAGATCGAGAGCTGCGGCGAGCACTGCTGCGTGGGCTTCAACACCC	4279
QY	1316	CCGACAAGACACAGAGGAGCCCCCTTCTGTGCCAT-----CGAGCTGCACCCCG	1369
DB	4280	CCGACAAGACACAGAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCAACCCCG	4339
QY	1370	ACAAAGTGGACCGTGACGCCCATCGAGCTGCCCGAGAGAGGAGCTGGAACCGTGAACGACA	1429
DB	4340	ACAAAGTGGACCGTGACGCCCATCGTGCCCGAGAGAGGAGCTGGAACCGTGAACGACA	4399
QY	1430	TCCAGAGCTGGTGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGATCAAGGTGC	1489
DB	4400	TCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGATCAAGGTGC	4459
QY	1490	GCCAGCTGTGCAAGCTGTGCGGCGCCCAAGGCCCTGACCGACATCTGTCGCCCTGACCCG	1549
DB	4460	GCCAGCTGTGCAAGCTGTGCGGCGCCCAAGGCCCTGACCGAGTGGTGTGCCCTGACCCG	4519
QY	1550	AGGAGCGCGAGCTGGAGCTGGCCCGAGAACCGCGAGATCTCTCGCGAGCCCGTGCACGGCG	1609
DB	4520	AGGAGCGCGAGCTGGAGCTGGCCCGAGAACCGCGAGATCTCTGAGGAGCCCGTGCACGGCG	4579
QY	1610	TGTTACTAGACCCAGCAAGGACCTGTGTGCGCGAGATCTCCAGAGAGCAGGGGCCAAGCAGT	1669
DB	4580	TGTTACTAGACCCAGCAAGGACCTGTGTGCGCGAGATCTCCAGAGAGCAGGGGCCAAGCAGT	4639

Qy	1670	GGACCTTACAGATCTTACGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGA	1720
Db	4640	GGACCTTACAGATCTTACGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCCGCA	4699
Qy	1730	TGCGCACCGGCCACACCAACGAGCTGAACAGCTGACCGAGGCGGTGCGAAGATCGCCA	1789
Db	4700	TGAAGGGCGGCCACACCAACGAGCTGAACAGCTGACCGAGGCGGTGCGAAGATCGCCA	4759
Qy	1790	TGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTTCGCTGCCCATCCAGAAGGAGA	1849
Db	4760	CCGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTTCAGCTGCCCATCCAGAAGGAGA	4819
Qy	1850	CCTGGGAGACCTTGGTGGACCGACTTCTGGCAGGCCACCTGGATCTCCCGAGTGGGAGTTCG	1909
Db	4820	CCTGGGAGGCTTGGTGGACCGACTTCTGGCAGGCCACCTGGATCTCCCGAGTGGGAGTTCG	4879
Qy	1910	TGAACACCCCCCTGGTGAAGCTGTGGTACCACTGGAGAGGAGCCCATCATCGGCG	1969
Db	4880	TGAACACCCCCCTGGTGAAGCTGTGGTACCACTGGAGAGGAGCCCATCATCGGCG	4939
Qy	1970	CCGAGACCTTCTAGCTGGACGGCGGCCCAACCGCGAGACCAAGATCGGCCAAGCCCGGCT	2029
Db	4940	CCGAGACCTTCTAGCTGGACGGCGGCCCAACCGCGAGACCAAGTGGGCAAGCCCGGCT	4999
Qy	2030	ACGTGACCGACCGGGCGCGCAGAAGATCTGTGACCTGACCCGAGACACCAACCAAGAAGA	2089
Db	5000	ACGTGACCGACCGGGCGCGCAGAAGTGGTGGCTTACCCGACACCAACCAAGAAGA	5059
Qy	2090	CCGAGCTGACAGGCATCTCAGCTGGCCCTTGACAGACAGCGGCACGAGGTGAACATCGTGA	2149
Db	5060	CCGAGCTGACAGGCATCTCAGCTGGCCCTTGACAGACAGCGGCCTGGAGGTGAACATCGTGA	5119
Qy	2150	CCGACACGAGTACGCCCTGGGCATCATCCAGGCCAGCCCGACAAAGAGCGAGCGAGCG	2209
Db	5120	CCGACACGAGTACGCCCTGGGCATCATCCAGGCCAGCCCGACAAAGAGCGAGCGAGCG	5179
Qy	2210	TGTTGAACACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGC	2269
Db	5180	TGTTGAGCCAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGC	5239
Qy	2270	CCGCCCAAGAGGCGATCGCGGGCAACAGACAGATTCGACAAGCTGGTGGAGCAAGGGCATCC	2329
Db	5240	CCGCCCAAGAGGCGATCGCGGGCAACAGACAGGTGGAACCGCTGGTGGAGCGCGGCATCC	5299
Qy	2330	GCAAGGTGCTTCTTCTGGAGCGCATCGATGGCGGCATCTGTGATCTACCACTA	2381
Db	5300	GCAAGGTGCTTCTTCTGGAGCGCATCGAAGGCCCGAGGAGACAGAA	5351
RESULT 11			
BD263706			
LOCUS	BD263706	2312 bp	DNA linear PAT 17-JUL-2003
DEFINITION	Improved expression of HIV polypeptides and production of virus-like particles.		
ACCESSION	BD263706		
VERSION	BD263706.1	GI:33073474	
KEYWORDS	JP 2002533124-A/73.		
SOURCE	synthetic construct		
ORGANISM	synthetic construct		
REFERENCE	1 (bases 1 to 2312)		
AUTHORS	Barnett, S., Megede, J.Z., Srivastava, I., Lian, Y., Hartog, K., Liu, H., Greer, C., Selby, M. and Walker, C.		
TITLE	Improved expression of HIV polypeptides and production of virus-like particles		
JOURNAL	Patent: JP 2002533124-A 73	08-OCT-2002;	
COMMENT	CHIRON CORP		
	OS Artificial Sequence		
	PN JP 2002533124-A/73		
	PD 08-OCT-2002		
	PF 30-DEC-1999	JP 2000591193	
	PR 31-DEC-1998	US 60/114495, 01-DEC-1999	US 60/168471
	SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI		

KARIN HARTOG,  
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC  
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC  
C12N5/10,  
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC  
Description of Artificial Sequence: FS(-). protmod. Rtopt (+) FH Key

FT source 1. .2312  
FT Location/Qualifiers  
FEATURES  
source Location/Qualifiers  
1. .2312  
/organism="Artificial Sequence".  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

## ORIGIN

Query Match 81.3%; Score 1998.4; DB 6; Length 2312;  
Best Local Similarity 92.9%; Pred. No. 1e-199;  
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;  
Qy 170 GCGCAAGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCAACTTCTTCGCG 229  
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCCTGAGAGACAGGCTAATTTCTTCGCG 60  
Qy 230 AGNACCTGGCTTCCCGCAGGCGAAGGCGCGAGTTCCCGAGGAGGAAACGCGCCA 289  
Db 61 AGNACCTGGCTTCCCGCAGGCGAAGGCGCGAGTTCCCGAGGAGGAAACGCGCCA 120  
Qy 290 ACAGCCCCACCGCGGAGTGCAGGTGGCGCGG-----ACAACCCCGGAGCGAGG 343  
Db 121 ACAGCCCCACCGCGGAGTGCAGGTGGCGCGGAGGAAACAGGCTGAGCGAGG 180  
Qy 344 CCGCGCGCGAGCGCGGACCCCTG-----AACTTCCCGCAGATCACTCTGGCGAGC 397  
Db 181 CCGCGCGCGAGCGCGGACCCGTGAGCTTCACTTCCCGCAGATCACTCTGGCGAGC 240  
Qy 398 GCGGCTGGTGGATCAAGGTGGCGCGCAGATCAAGAGGCGCTCTGCGACCGCGG 457  
Db 241 GCGGCTGGTGGATCAAGGTGGCGCGCAGTCAAGAGGCGCTCTGCGACCGCGG 300  
Qy 458 CCGACGACCGTCTCGAGGAGATGAGCTGCGCGGAGTGGAGCGGAGCGGAGCGGATG 517  
Db 301 CCGACGACCGTCTCGAGGAGATGAACTTGGCGCGGAGTGGAGCGGAGCGGATG 360  
Qy 518 GCGGATCGGCGGCTTCATCAAGGTGGCGCGACAGTACGACAGATCCTCATCGAGATCTGCG 577  
Db 361 GCGGATCGGCGGCTTCATCAAGGTGGCGCGACAGTACGACAGATCCTCATCGAGATCTGCG 420  
Qy 578 GCAAGAGGCGCATCGGCACCGTGTGATCGGCGCGGCGGCGGCGGCGGCGGCGGCGCA 637  
Db 421 GCGCAAGGCGCATCGGCACCGTGTGATCGGCGCGGCGGCGGCGGCGGCGGCGGCGCA 480  
Qy 638 ACATGCTGACCGAGTGGGCTGACCTTGAACCTTCCCGCATCGGCGGCGGCGGCGGCGG 697  
Db 481 ACCTGCTGACCGAGTGGGCTGACCTTGAACCTTCCCGCATCGGCGGCGGCGGCGGCGG 540  
Qy 698 CCGTGAAGCTGAAGCGCGGCGATGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 757  
Db 541 CCGTGAAGCTGAAGCGCGGCGATGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
Qy 758 AGAAGATCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 817  
Db 601 AGAAGATCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
Qy 818 AGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 877  
Db 661 AGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720  
Qy 878 CCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 937  
Db 721 CCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
Qy 938 AGGTGACGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 997

Db 781 AGGTGACGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
Qy 998 TGGACGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1057  
Db 841 TGGACGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
Qy 1058 CTTTACCATCTCCCGAGCATCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117  
Db 901 CTTTACCATCTCCCGAGCATCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Qy 1118 TGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1177  
Db 961 TGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
Qy 1178 AGCCCTTTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1231  
Db 1021 AGCCCTTTCGCGGAGGAG 1080  
Qy 1232 TGGGCGAGCGACCTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1291  
Db 1081 TGGGCGAGCGACCTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
Qy 1292 TGGTGGCTGGGCGGCTTCAACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1351  
Db 1141 TGGTGGCTGGGCGGCTTCAACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
Qy 1352 CCAT-----CGAGCTGCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1405  
Db 1201 GGATGGGCTACGAGCTGCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
Qy 1406 AGGAGAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1465  
Db 1261 AGGAGAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320  
Qy 1466 AGATCTTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1525  
Db 1321 AGATCTTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
Qy 1526 TGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1585  
Db 1381 TGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440  
Qy 1586 TCCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1645  
Db 1441 TCCTGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500  
Qy 1646 TCAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1705  
Db 1501 TCAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560  
Qy 1706 TGAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1765  
Db 1561 TGAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620  
Qy 1766 CCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1825  
Db 1621 CCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
Qy 1826 TCAGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1885  
Db 1681 TCAAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740  
Qy 1886 CTTGGATCCCGGAGTGGGAGTTCGTGAAACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1945  
Db 1741 CTTGGATCCCGGAGTGGGAGTTCGTGAAACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800  
Qy 1946 TGAAGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2005  
Db 1801 TGAAGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860  
Qy 2006 AGACCAAGATCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2065





```
|||||
1261 AGGACAGTGGACCGTGAACACATCCAGAGCTGGTGGCAAGCTGAATCGGCCAGCC 1320
Qy |||||
1466 AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTCAAGCTGCTCGCGGCGCCCAAGGCC 1525
Db |||||
1321 AGATCTAGCGCGGCATCAAGGTGAAGCAGCTGTCAAGCTGCTCGCGGCGCCCAAGGCC 1380
Qy |||||
1526 TGACCGACATCTGTCCCTTGACCGAGGAGCGCAGCTGGAGCTGGCCGAGAACCGCAGGA 1585
Db |||||
1381 TGACCGAGGTGATCCCTTGACCGAGGAGCGCAGCTGGAGCTGGCCGAGAACCGCAGGA 1440
Qy |||||
1586 TCCTGCGGAGCCCGTGTGACCGGCTGTACTAGACCCCGCAGGACCTGTGTGCCGAGA 1645
Db |||||
1441 TCCTGAAGAGCCCGTGTGACCGAGGTGACTACGACCCCGCAGGAGCCTGTGTGCCGAGA 1500
Qy |||||
1646 TCCAGAGAGCGGCGCACGACCAAGTGGACCTACACAGATCTACACGAGCGCTTCAAGAAC 1705
Db |||||
1501 TCCAGAGAGCGGCGCACGAGGCGCAGTGGACCTACACAGATCTACACGAGCGCTTCAAGAAC 1560
Qy |||||
1706 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGACGTGAAGCAGCTGA 1765
Db |||||
1561 TGAAGACCGGCAAGTACGCCGATGCGCGCGCCACACCAACGACGTGAAGCAGCTGA 1620
Qy |||||
1766 CCGAGGCGGTGCAGAAATCGCCATGAGAGACATCGTGATCTGGGGCAAGACCCCAAGT 1825
Db |||||
1621 CCGAGGCGGTGCAGAAATCGCCATGAGAGACATCGTGATCTGGGGCAAGATCCCAAGT 1680
Qy |||||
1826 TCCGCTTCCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTACTTGGCAGGCCA 1885
Db |||||
1681 TCAAGCTGCCATCCAGAGAGACCTGGGAGGCTGTGTGATGAGTACTTGGCAGGCCA 1740
Qy |||||
1886 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTGTACAGC 1945
Db |||||
1741 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTGTACAGC 1800
Qy |||||
1946 TGGAGAGAGGCCCATCATCTGGCGCCGAGACCTTCTAGTGGAGCGGCGCGGCCAACCGCG 2005
Db |||||
1801 TGGAGAGAGGCCCATCTGTGGCGCCGAGACCTTCTAGTGGAGCGGCGCGGCCAACCGCG 1860
Qy |||||
2006 AGACCAAGATCGCAAGCGCGCTACGTGACCGACCGGCGCGGCAGCAAGATCGTAGCC 2065
Db |||||
1861 AGACCAAGTGGGCAAGCGCGCTACGTGACCGACCGGCGCGGCAGCAAGTGTGTAGCA 1920
Qy |||||
2066 TGACCGAGACACCAACAGAGACCGAGCTGACAGGCCATCCAGCTGGCCCTGCAGGACA 2125
Db |||||
1921 TCGCCGACACCAACAGAGACCGAGCTGACAGGCCATCCACCTGGCCCTGCAGGACA 1980
Qy |||||
2126 CGCGCAGCGAGTGAAATCTGTGACCGACAGCCAGTACGCCCTGGGCAATATCCAGGCC 2185
Db |||||
1981 CGCGCTTGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCAATATCCAGGCC 2040
Qy |||||
2186 AGCCCGACAGAGCGAGCGAGCTGTGAACACAGATCATCGACGACGTATCAAGAGG 2245
Db |||||
2041 AGCCCGACAGAGCGAGCGAGCTGTGTGAGCCAGATCATCGACGACGTATCAAGAGG 2100
Qy |||||
2246 AGAAGTGTACTCTAGCTGGGTGCGCCCGCCACAGAGGCGCATCGCGGCAAGCAGATCG 2305
Db |||||
2101 AGAAGTGTACTCTGGCTGGGTGCGCCCGCCACAGAGGCGCATCGCGGCAAGCAGTGG 2160
Qy |||||
2306 ACAAGCTGTGAGCAAGGCAATCCGCAAGGTGTCTTCTGAGAGCGGCAATCGATGGCGCA 2365
Db |||||
2161 ACAAGCTGTGAGCGCGGCGCATCCGCAAGGTGTCTTCTGAGAGCGGCAATCGATGGCGCA 2220
Qy |||||
2366 TCGTGATCTACAGTACATGACGACCTGTACGTGGGCGCGGCGCCCTAGGATCGATT 2425
Db |||||
2221 TCGTGATCTACAGTACATGACGACCTGTACGTGGGCGCGGCGCCCTAGGATCGATT 2280
Qy |||||
2426 AAAAGCTTCCCGGCGCTAGCACCGGTGAATTC 2457
Db |||||
2281 AAAAGCTTCCCGGCGCTAGCACCGGTGAATTC 2312
```

RESULT 13

```
AR373389
LOCUS AR373389 2312 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 84 from patent US 6602705.
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like
particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
FEATURES
source
1. 2312
/mol_type="genomic DNA"

ORIGIN
Query Match 81.3%; Score 1998.4; DB 6; Length 2312;
Best Local Similarity 92.9%; Pred. No. 1e-199;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

Qy 170 GCGGCAAGAGGCGCCACCATGAAGACTGACACGAGCGCCAGGCGCAACTTCTTCGCG 229
Db 1 GCGGCGCGGAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAAATTTCTTCGCG 60

Qy 230 AGNACCTGGCTTCCCCCAGGCGCAAGCCCGCGAGTTCCCCAGGAGCAGACCGCGCA 289
Db 61 AGNACCTGGCTTCTCTCAGGCGCAAGCCCGCGAGTTTACGAGCGAGCAGACCGCGCCA 120

Qy 290 ACAGCCCCACAGCGCGAGCTGCAGGTGCGCGCGG-----ACAAACCCCGCAGCGAGG 343
Db 121 ACAGCCCCACCGCGCGAGCTGCAGGTGTGGGCGCGGAGAACACAGCTGTGAGCGAGG 180

Qy 344 CCGGCGCGAGCGCCAGGCGACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGC 397
Db 181 CCGGCGCGAGCGCCAGGCGACCGTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC 240

Qy 398 GCCCTCTGTGAGCATCAAGTGGGCGCGCAGATCAAGGAGGCCCTCTGTGACACCGCGCG 457
Db 241 GCCCTCTGTGTGACCATCAGGATCGGCGCGCAGCTCAAGGAGGCGCTCTGTGACACCGCGCG 300

Qy 458 CCGACGACACCGCTGTGAGGAGATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGACGACACCGTGTGAGGAGATGAACTGCCCGGCAAGTGAAGCCCAAGATGATCG 360

Qy 518 GCGGCATCGGCGGCTTTCATCAAGGTGCGCCAGTACGACCAAGATCCTGATCGAGATCTGCG 577
Db 361 GCGGCATCGGCGGCTTTCATCAAGGTGCGGCGAGTACGACCAAGATCCTGATCGAGATCTGCG 420

Qy 578 GCAAGAGGCGCATCGGCACCGTGTGATCGGCGCGCACCCCGGTGAACATCATCGGCGCGCA 637
Db 421 GCAAGAGGCGCATCGGCACCGTGTGATCGGCGCGCACCCCGGTGAACATCATCGGCGCGCA 480

Qy 638 ACATGCTGACCGAGCTGGGCTGCACCTGAACTTCCCCCATCAGCCCATCGAGACCCTGCG 697
Db 481 ACCTGCTGACCGAGATCGGCTGCACCTGAACTTCCCCCATCAGCCCATCGAGACCCTGCG 540

Qy 698 CCGTGAAGCTGAAGCCCGGATGGAACCGCCCGCAAGGTGAAGAGTGGCCCTGTGACCGAGG 757
Db 541 CCGTGAAGCTGAAGCCCGGATGGAACCGCCCGCAAGGTGAAGAGTGGCCCTGTGACCGAGG 600

Qy 758 AGAAGATCAAGGCGCTGACCGCCATCTGCAGGAGATGGAAGAGGAGGCGGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCGCTGAGGAGATCTGCACCGGAGATGGAAGAGGAGGCGGCAAGATCACCA 660

Qy 818 AGATCGGCGCGGAGAACCCCTTACAACACCCCGTGTTCGCGCATCAAGAGAGGAGCAGCA 877
Db 661 AGATCGGCGCGGAGAACCCCTTACAACACCCCGTGTTCGCGCATCAAGAGAGGAGCAGCA 720

Qy 878 CCAAGTGGCGCAAGCTGGTGGACTTTCGCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGG 937
```



Db 3081 ACTGCGGCGCCCGGCAAGAGAGGCTGCTGGAAGTGGGCAAGAGAGGCGCACAGGTGA 3140  
Qy 194 AGGACTGCACGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGGCA 253  
Db 3141 AGGACTGCACGAGCGGACAGGCTAA-----TAGGGAAGATCTGGCCTTCCCAAGGGA 3194  
Qy 254 AGGCGCGGAGTTTCCAGCGAGAGAGAAACCGCGCAACAGGCCCAACAGCGCGAGCTGC 313  
Db 3195 AGGCGAGGGAATTTCTTACAGAGCAGACCAAGAGCAACAGCCCAACAGAGAGAGCTTC 3254  
Qy 314 AGGTGCGCGG-----CGACAACCCCGCAGCGAGGCGGCGCGAGCGCGCAGGCA--- 364  
Db 3255 AGGTTTGGGAGAGACAACTCTCTCAGAGAGAGAGCGCCGATAGACAAGAACTG 3314  
Qy 365 ---CCCTGAACTTCCCGCAGATCACTCTGTGGCAGCGCCCGCTGGTGAAGCATCAAGTGG 421  
Db 3315 TATCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACATAAGATAG 3374  
Qy 422 GCGGCCAGATCAGGAGGCGCTGTGGACACCGCGCGCGAGGACACCGTGTGAGAGAGA 481  
Db 3375 GGGGCCAGCTGAAGAGGCGCTTCTAGACACCGCGCGCGAGCAACCGTGTGAGAGAGA 3434  
Qy 482 TGAGCCTGCCCGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTCAATCAAG 541  
Db 3435 TGAACCTGCCCGCGCTGGAAGCCCAAGATGATCGCGGCGATCGCGGCTTCAATCAAG 3494  
Qy 542 TGCGCCAGTAGACCAAGATCTCGATCGAGATCTCGGCAAGAGGCCATCGGCAACCGTGC 601  
Db 3495 TGGGCCAGTAGACCAAGATCTCGATCGAGATCTCGGCGCACAGGCCATCGGCAACCGTGC 3554  
Qy 602 TGATCGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGAACACAGCTGGGTGCA 661  
Db 3555 TGGTGGCGCCACCCCGTGAACATCATCGCGCGCAACCTGCTGACCCAGATCGGCTGCA 3614  
Qy 662 CCCTGAACTTCCCATCAGCCCATCGAGCCGTCAGACCGTGCAGCTGAAGCCCGGATGG 721  
Db 3615 CCCTGAACTTCCCATCAGCCCATCGAGCCGTCAGACCGTGCAGCTGAAGCCCGGATGG 3674  
Qy 722 ACGCCCAAGAGTGAAGCAGTGGCCCTCGACCGAGAGAGATCAAGCCCTGACCGCA 781  
Db 3675 ACGCCCAAGAGTGAAGCAGTGGCCCTCGACCGAGAGAGATCAAGCCCTGCTGTGAGA 3734  
Qy 782 TCTGCGAGGAGATGGAGAGAGGCAAGATCAACAAAGATCGGCGCGAGAACCCCTTACA 841  
Db 3735 TCTGCAACGAGATGGAGAGAGGCGCAAGATCAGCAAGATCGGCGCGAGAACCCCTTACA 3794  
Qy 842 ACACCCCGTGTTCGCCATCAAGNAGAGAGACAGCAACAGTGGCGCAAGCTGTGACT 901  
Db 3795 ACACCCCGTGTTCGCCATCAAGNAGAGAGACAGCAACAGTGGCGCAAGCTGTGACT 3854  
Qy 902 TCGCGAGCTGAACAAGCGCACCCAGGACTTCTTGGGAGGTGACAGTGGGCATCCCCACC 961  
Db 3855 TCGCGAGCTGAACAAGCGCACCCAGGACTTCTTGGGAGGTGACAGTGGGCATCCCCACC 3914  
Qy 962 CCGCGGCTCTGAAGAAGAGAGCGTGACCGTGTGAGCGCTGGGCGAGCGCTACTTCA 1021  
Db 3915 CCGCGGCTCTGAAGAAGAGAGCGTGACCGTGTGAGCGCTGGGCGAGCGCTACTTCA 3974  
Qy 1022 CGGTGCCCTTGAGCAGGACTTTCGCAAGTACACCGCTTCAACATCCCGCAGAGTCAACA 1081  
Db 3975 GCGTGCCCTTGAGCAAGACTTTCGCAAGTACACCGCTTCAACATCCCGCAGAGTCAACA 4034  
Qy 1082 ACAGAGCCCGCGCATCCGCTACAGTACACCGTGTGCGCCAGGCTGGAAGGGCAGCC 1141  
Db 4035 ACAGAGCCCGCGCATCCGCTACAGTACACCGTGTGCGCCAGGCTGGAAGGGCAGCC 4094  
Qy 1142 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCCGCGCCCGCAACCCCG 1201  
Db 4095 CCGCATCTTCCAGTGCAGCATGACCAAGATCTCTGGAGCCCTTCCCGCAAGCAAGACCCCG 4154  
Qy 1202 AGATCGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGCATCTGGAGATCGGCC 1255  
Db 4155 ACATCGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGCATCTGGAGATCGGCC 4214

Qy 1256 AGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTCCACCCC 1315  
Db 4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGTGGGGCTTCCACCCC 4274  
Qy 1316 CCGCAAGAAGACCAAGAGAGGCCCCCTTCTTCCCTGCCCCAT-----CGAGTGCACCCCG 1369  
Db 4275 CCGCAAGAAGACCAAGAGAGGCCCCCTTCTTCTGTGGATGGGCTACGAGTGCACCCCG 4334  
Qy 1370 ACAAGTGCACCGTGCAGCCCATCGAGCTGCGCGAAGAGAGAGCTGACCGTGAACGACA 1429  
Db 4335 ACAAGTGCACCGTGCAGCCCATCGTCTGCGCGAAGAGAGAGCTGACCGTGAACGACA 4394  
Qy 1430 TCAGAAAGCTGTGGGCAAGCTGAACTGGGCGACGCCAGATCTTACCCCGCATCAAGGTGC 1489  
Db 4395 TCAGAAAGCTGTGGGCAAGCTGAACTGGGCGACGCCAGATCTTACCGCGCATCAAGGTGC 4454  
Qy 1490 GCAGCTGTGCAAGCTGCTGCGCGCGCAAGGCGCTGACCGCATCTGTCGCCCTGACCG 1549  
Db 4455 GCAGCTGTGCAAGCTGCTGCGCGCGCAACGAGGCGCTGACCGAGGTGTGCCCCGACCG 4514  
Qy 1550 AGAGGCGCAGCTGGAGCTGCGCGAGAAACCGAGATCTTGCAGGAGCCCGTGCACGCGC 1609  
Db 4515 AGAGGCGCAGCTGGAGCTGCGCGAGAAACCGAGATCTTGCAGGAGCCCGTGCACGCGC 4574  
Qy 1610 TGTACTACGACCCCGCAGCAAGGACCTGTGTGCGCGAGATCCAGAAAGCAGGCGCACGACGAT 1669  
Db 4575 TGTACTACGACCCCGCAGCAAGGACCTGTATCGCGAGATCCAGAAAGCAGGCGCACGACGAT 4634  
Qy 1670 GGAACCTACAGATCTACAGAGCGCTTCAAGAACCTGAAAGACCGCGAAGTACCGCGCA 1729  
Db 4635 GGAACCTACAGATCTACAGAGCGCTTCAAGAACCTGAAAGACCGCGAAGTACCGCGCA 4694  
Qy 1730 TGGCGCACCGCCACACACGAGCTGAGCAGCTGACCGAGGCGCTGACGAGATCGCCA 1789  
Db 4695 TGAAGGCGCGCCACACACGAGCTGAGCAGCTGACCGAGGCGCTGACGAGATCGCCA 4754  
Qy 1790 TGGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTTCCGCTTCCGCTTCCGCAAGGAGA 1849  
Db 4755 CCGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTTCAAGTTCAGCTGCCCATCCAGAGGAGA 4814  
Qy 1850 CTTGGAGAGCTCTGGTGCACCGACTATCTGGCAGGCGACCTGGATCCCGAGTGGAGTTG 1909  
Db 4815 CTTGGAGGCTCTGGTGCACCGACTATCTGGCAGGCGACCTGGATCCCGAGTGGAGTTG 4874  
Qy 1910 TGAACACCCCGCTGTTGAGCTGTGTTACAGCTGAGAGAGGAGCCCATCATCGCGC 1969  
Db 4875 TGAACACCCCGCTGTTGAGCTGTGTTACAGCTGAGAGAGGAGCCCATCATCGCGC 4934  
Qy 1970 CCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGCT 2029  
Db 4935 CCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGTGGGCGAGGCCGCT 4994  
Qy 2030 ACCTGACCGAGCGCGCGCAGAAAGATCTGTAGCTTGCAGACCAACCAACAGAGAGA 2089  
Db 4995 ACCTGACCGAGCGCGCGCAGAAAGTGTGTGCTGCTGACCGCACCAACCAACAGAGAGA 5054  
Qy 2090 CCGAGCTGAGGCGCATCCAGCTGGCCCTGACAGNACAGCGGCGAGGAGTGAACATCTGTA 2149  
Db 5055 CCGAGCTGAGGCGCATCCACTTGGCCCTGACAGNACAGCGGCGAGGAGTGAACATCTGTA 5114  
Qy 2150 CCGAGCAGCATGCGCCCTGGGCGATCATCCAGCGCCAGCGCCCAACAGAGCGAGAGCGAGC 2209  
Db 5115 CCGAGCAGCATGAGCGCCCTGGGCGATCATCCAGCGCCAGCGCCCAACAGAGCGAGAGCGAGC 5174  
Qy 2210 TGGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACTCTGAGCTGGGTGC 2269  
Db 5175 TGGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACTCTGAGCTGGGTGC 5234  
Qy 2270 CCGCCCAAGAGGCGATCGGCGCAACGAGCAGATTCGACAGCTGCTGAGCAGAGGCGATCC 2329  
Db 5235 CCGCCCAAGAGGCGATCGGCGCAACGAGCAGGTGGAACGCGCTTGGTGAAGCGCGCATCC 5294



```
||||| 4643 GGACCTACAGATCTACAGAGGCCCTTCAAGAACCTGAAGACCGGAAGTACGCCCGCA 4702
||||| 1730 TGCACACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCGGTGCAGAAAGATCGCCA 1789
||||| 4703 TGAAGGGCGCCCAACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCA 4762
||||| 1790 TGAAGAGATCGTGATCTGGGGCAAGACCCCAAGTTCCGGCTGCCCATCCAGAAGGAGA 1849
||||| 4763 CCGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCAAGCTGCCATCCAGAAGGAGA 4822
||||| 1850 CCTGGGAGACCTGTGACCGACTACTTGGCAGAGGCCACTGGATCCCGAGTGGGAGTTGG 1909
||||| 4823 CCTGGGAGGCTGTGTGACCGAGTACTTGGCAGAGGCCACTTGGATCCCGAGTGGGAGTTGG 4882
||||| 1910 TGAACACCCCCCTCGTGAAGCTGTGGTACCAAGCTGGGAAGAGGCCCATCATCGGCG 1969
||||| 4883 TGAACACCCCCCTCGTGAAGCTGTGGTACCAAGCTGGGAAGAGGCCCATCATCGGCG 4942
||||| 1970 CCGAGACCTTCTACGTGGAAGCGCCCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
||||| 4943 CCGAGACCTTCTACGTGGAAGCGCCCGCCCAACCGCGAGACCAAGTGGGCAAGGCCGGCT 5002
||||| 2030 ACGTGACCGACCGGGCGCGCAGAGATCGTGAGCCTGACCGAGACCAACCAACCAAGA 2089
||||| 5003 ACGTGACCGACCGGGCGCGCAGAGAGTGGTGGCCCTGACCGAGACCAACCAACCAAGA 5062
||||| 2090 CCGAGCTGCAGGCCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGGTGAACATCGTGA 2149
||||| 5063 CCGAGCTGCAGGCCCATCCAGCTGGCCCTGCAGGACAGCGGCGCTGGAGGTGAACATCGTGA 5122
||||| 2150 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGCCAGCCGACAGAGCGAGAGCGAGC 2209
||||| 5123 CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGCCAGCCGACAGAGCGAGAGCGAGC 5182
||||| 2210 TGGTGAACCCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGTAGCTGGGTGC 2269
||||| 5183 TGGTGAACCCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGTAGCTGGGTGC 5242
||||| 2270 CCGCCCAACAGGGGCATCGGGGCAACGAGCAGATCGACCAAGCTGGTGAAGCAAGGGCATCC 2329
||||| 5243 CCGCCCAACAGGGGCATCGGGGCAACGAGCAGGTGGACGGCCTGGTGAAGCGCGGCATCC 5302
||||| 2330 GCAAGGTGCTTCTCGGACGGGCATCGATGGCGGCATCGTGATCTACAGTA 2381
||||| 5303 GCAAGGTGCTTCTCGGACGGGCATCGACCAAGGCCCGAGGAGGACGACGAGAA 5354
```

Search completed: June 1, 2005, 20:20:49  
Job time : 10525 secs



**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2005, 23:20:00 ; Search time 1264.25 Seconds  
(without alignments)  
11504.726 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457

Sequence: 1 gtcagccaccatggccga.....gggctagaccggatgaattc 2457

Scoring table: IDENTITY NUC

Gapop 10\*0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	100.0	2457	6	ABL39961
2	2457	100.0	2457	12	ADM73766
3	2443.4	99.4	2445	8	ACA03546
4	2443.4	99.4	2445	10	ADCI3264
5	2436.4	99.2	3930	10	ADCI3230
6	2436.2	99.2	2463	6	ABL39960
7	2436.2	99.2	2463	12	ADM73765
8	2434.8	99.1	3930	10	ADCI3231
9	2434.8	99.1	3930	10	ADCI3232
10	2434.8	99.1	5184	8	ACA03591
11	2434.8	99.1	5184	10	ADCI3279
12	2428.6	98.8	2457	8	ACA03547
13	2428.6	98.8	2457	10	ADCI3265
14	2415.4	98.3	2469	6	ABL39959
15	2415.4	98.3	2469	12	ADM73764
16	2404.4	97.9	3531	8	ADCI3234
17	2403.4	97.8	2457	8	ACA03548
18	2403.4	97.8	2457	10	ADCI3266
19	2402.8	97.8	3537	10	ADCI3236
20	2401.8	97.8	5145	8	ACA03521

21	2401.8	97.8	5145	10	ADCI3233
22	2391.8	97.3	3538	10	ADCI3235
23	2343.4	95.4	3607	8	ACA03551
24	2343.4	95.4	3607	10	ADCI3269
25	2325.2	94.6	3597	8	ACA03549
26	2325.2	94.6	3597	10	ADCI3267
27	2295.8	93.4	3624	8	ACA03550
28	2295.8	93.4	3624	10	ADCI3268
29	2136	86.9	2460	8	ACA03541
30	2136	86.9	2460	8	ACC78505
31	2135	86.9	3564	8	ACC78488
32	2135	86.9	3564	8	ACC78489
33	2134.4	86.9	4716	10	ADCI3238
34	2134.4	86.9	4716	10	ADCI3238
35	2130.6	86.7	3999	8	ACC78484
36	2129	86.7	3999	8	ACC78485
37	2129	86.7	3999	8	ACC78486
38	2129	86.7	5283	8	ACA03584
39	2129	86.7	5283	8	ACC78529
40	2128.8	86.6	4713	8	ACA03592
41	2128.8	86.6	4713	10	ADCI3280
42	2128.4	86.6	3462	10	ADCI3237
43	2115.2	86.1	2466	8	ACA03542
44	2115.2	86.1	2466	8	ACC78506
45	2103	85.6	3735	8	ACA03545

ALIGNMENTS

RESULT 1

ABL39961

ID ABL39961 standard; DNA; 2457 BP.

XX

AC ABL39961;

XX

DT 15-MAY-2002 (first entry)

XX

DE Synthetic construct PR975YMM SEQ ID NO:32.

XX

KW Human immunodeficiency virus type C; antigenic HIV type C protein; immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; nef; immunostimulant; gene therapy; gene; ds.

XX

OS Human immunodeficiency virus; type C.

OS Synthetic.

PN WO200204493-A2.

XX

PD 17-JAN-2002.

XX

PF 05-JUL-2001; 2001WO-US021241.

XX

PR 05-JUL-2000; 2000US-00610313.

XX

PA (CHIR ) CHIRON CORP.

PA (UYST-) UNIV STELLENBOSCH.

XX

PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;

XX

DR WPI; 2002-154920/20.

XX

PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.

PT

XX Claim 1; Fig 10; 233pp; English.

XX

CC The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef (I). (I) have immunostimulant activity and can be used in gene therapy.

CC

CC The HIV type C polynucleotides are useful in applications including DNA  
 CC immunisation, generation of packaging cell lines, and production of HIV  
 CC Type C proteins. The polynucleotides are particularly useful in gene  
 CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
 CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
 CC the present invention  
 XX

SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match		100.0%; Score 2457; DB 6; Length 2457;
Best Local Similarity		100.0%; Pred. No. 9.7e-296;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	GTGCGAGCCACCATGCGGAGCCATGAGCCAGGCGACAGCGCCAAATCTGTATGCAG 60
DB	1	GTGCGAGCCACCATGCGGAGCCATGAGCCAGGCGACAGCGCCAAATCTGTATGCAG 60
QY	61	CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGAGGCG 120
DB	61	CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGAGGCG 120
QY	121	CACATGCGCGCAACTGCGCGCCCTCCCGCAAGAAAGGGCTGTGAAATGCGGCAAGGAG 180
DB	121	CACATGCGCGCAACTGCGCGCCCTCCCGCAAGAAAGGGCTGTGAAATGCGGCAAGGAG 180
QY	181	GGCCACCAAGATGAAGGATGCAACGAGCGCCAGGCGCAACTTCTTCCGCGAGGACCTGGCC 240
DB	181	GGCCACCAAGATGAAGGATGCAACGAGCGCCAGGCGCAACTTCTTCCGCGAGGACCTGGCC 240
QY	241	TTTCCCGCAGGGCAAGGCCCGGAGTTCCCGAGCGAGCAAGCCGCGCAAGCGCCACCC 300
DB	241	TTTCCCGCAGGGCAAGGCCCGGAGTTCCCGAGCGAGCAAGCCGCGCAAGCGCCACCC 300
QY	301	AGCCGCGAGCTGCAAGTGGCGGCGCAAAACCCCGCAGCGAGGCGCGGCGCGAGCGCCAG 360
DB	301	AGCCGCGAGCTGCAAGTGGCGGCGCAAAACCCCGCAGCGAGGCGCGGCGCGAGCGCCAG 360
QY	361	GGCAGCTTGAATCTCCCGAGATCACTGTGGAGCGCCCTTGTGAGGAGATCAAGGTG 420
DB	361	GGCAGCTTGAATCTCCCGAGATCACTGTGGAGCGCCCTTGTGAGGAGATCAAGGTG 420
QY	421	GGCGGCCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCAACCGTGTCTGAGGAG 480
DB	421	GGCGGCCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCAACCGTGTCTGAGGAG 480
QY	481	ATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCAAG 540
DB	481	ATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCAAG 540
QY	541	GTGCGCCAGTACGACAGATCTGTGATCGAGATCTGCGGCGAGAGGCGCATCGGCACCGTG 600
DB	541	GTGCGCCAGTACGACAGATCTGTGATCGAGATCTGCGGCGAGAGGCGCATCGGCACCGTG 600
QY	601	CTGATCGGCGCCCAACCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
DB	601	CTGATCGGCGCCCAACCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
QY	661	ACCTTGAATCTTCCCATCAGCCCCATCGAGACCGTGTGCGGTGAAGTGAAGCCCGGCATG 720
DB	661	ACCTTGAATCTTCCCATCAGCCCCATCGAGACCGTGTGCGGTGAAGTGAAGCCCGGCATG 720
QY	721	GACGGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780
DB	721	GACGGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780
QY	781	ATCTCGAGGAGATGAGAGAGGCGCAAGATCACCAAGATCGGCGCCGAGAACCCCTTAC 840
DB	781	ATCTCGAGGAGATGAGAGAGGCGCAAGATCACCAAGATCGGCGCCGAGAACCCCTTAC 840
QY	841	AACACCCCGTGTTCGCATCAAGAAAGAGGAGCAACAAGTGGCGCAAGCTGTGTGAC 900
DB	841	AACACCCCGTGTTCGCATCAAGAAAGAGGAGCAACAAGTGGCGCAAGCTGTGTGAC 900

QY	901	TTCCGCGAGCTGAACAACAGCGCACCCAGGACTTCTGGGAGGTGACAGTGGGCATCCCCAC 960
DB	901	TTCCGCGAGCTGAACAACAGCGCACCCAGGACTTCTGGGAGGTGACAGTGGGCATCCCCAC 960
QY	961	CCCGCGGCTTGAAGAAGAAGAGCGTGACCGTGTGGAGCGTGGGCGAGCGCTTACTTC 1020
DB	961	CCCGCGGCTTGAAGAAGAAGAGCGTGACCGTGTGGAGCGTGGGCGAGCGCTTACTTC 1020
QY	1021	AGCGTGCCCTTGGAGAGGACTTTCGCAAGTACACCGCTTTCACATCCCGAGATCAAC 1080
DB	1021	AGCGTGCCCTTGGAGAGGACTTTCGCAAGTACACCGCTTTCACATCCCGAGATCAAC 1080
QY	1081	AACGAGACCCCGCGCATCCGCTACCAAGTACCAAGTGTCTGCCAGGGTGGAGGGGAGC 1140
DB	1081	AACGAGACCCCGCGCATCCGCTACCAAGTACCAAGTGTCTGCCAGGGTGGAGGGGAGC 1140
QY	1141	CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
DB	1141	CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCC 1200
QY	1201	GAGATCGTGATCTTACAGGCCCGCTGTACGTGGGAGCGGACCTGGAGATCGGCCAGCAC 1260
DB	1201	GAGATCGTGATCTTACAGGCCCGCTGTACGTGGGAGCGGACCTGGAGATCGGCCAGCAC 1260
QY	1261	CGCGCAAGATCGAGGAGCTGCGAAGCACCTGTGCGCTGGGGCTTCAACACCCCGGAC 1320
DB	1261	CGCGCAAGATCGAGGAGCTGCGAAGCACCTGTGCGCTGGGGCTTCAACACCCCGGAC 1320
QY	1321	AAGAAGCACCAAGAGGAGCCCTTCTGCGCATTCGAGCTGCAACCCCGCAAGTGGAGCC 1380
DB	1321	AAGAAGCACCAAGAGGAGCCCTTCTGCGCATTCGAGCTGCAACCCCGCAAGTGGAGCC 1380
QY	1381	GTGAGCCCATTCGAGCTGCTCCGAGAGGAGAGCTGGACCTGTGAACGAGATCCAGAGCTG 1440
DB	1381	GTGAGCCCATTCGAGCTGCTCCGAGAGGAGAGCTGGACCTGTGAACGAGATCCAGAGCTG 1440
QY	1441	GTGGCAAGCTGAATCTGGGCGCAGCAGATCTACCCCGGCATCAAGTGTGCGCAGCTGTGC 1500
DB	1441	GTGGCAAGCTGAATCTGGGCGCAGCAGATCTACCCCGGCATCAAGTGTGCGCAGCTGTGC 1500
QY	1501	AAGTGTCTGCGCGCGCCCAAGGCCCTGTGACCGACATCTGTCGCCCTGACCGAGAGGCCAG 1560
DB	1501	AAGTGTCTGCGCGCGCCCAAGGCCCTGTGACCGACATCTGTCGCCCTGACCGAGAGGCCAG 1560
QY	1561	CTGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGTGACCGGCTGTACTACGAC 1620
DB	1561	CTGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGTGACCGGCTGTACTACGAC 1620
QY	1621	CCCAGAGGAGCTGCTGGCGCGAGATCCAGAAAGAGAGGCGCACAGCTGGAGCTTACCGAG 1680
DB	1621	CCCAGAGGAGCTGCTGGCGCGAGATCCAGAAAGAGAGGCGCACAGCTGGAGCTTACCGAG 1680
QY	1681	ATCTACAGGAGCCCTTCAAGAACTGAAGAACCGGCAAGTACGCGCAAGATGCGCACCGCC 1740
DB	1681	ATCTACAGGAGCCCTTCAAGAACTGAAGAACCGGCAAGTACGCGCAAGATGCGCACCGCC 1740
QY	1741	CACACCAAGAGCTGAAGCAGCTGACCGAGCGCTGAGAGATCGCCATGGAGAGCATC 1800
DB	1741	CACACCAAGAGCTGAAGCAGCTGACCGAGCGCTGAGAGATCGCCATGGAGAGCATC 1800
QY	1801	GTGATCTGGGCGCAAGACCCCAAGTTCCGCTTCCCATCCAGAAAGGAGACTCTGGGAGACC 1860
DB	1801	GTGATCTGGGCGCAAGACCCCAAGTTCCGCTTCCCATCCAGAAAGGAGACTCTGGGAGACC 1860
QY	1861	TGGTGGACCGACTACTGGCAGCCACCTGGATCTCCCGAGTGGAGTTCGTGAACACCCGCC 1920
DB	1861	TGGTGGACCGACTACTGGCAGCCACCTGGATCTCCCGAGTGGAGTTCGTGAACACCCGCC 1920
QY	1921	CCCTTGTGTGAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTC 1980
DB	1921	CCCTTGTGTGAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTC 1980
QY	1981	TACGTGGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGCGGCTACGTGACCGGAC 2040

Db 1981 TACGTGGACGGCGCGCCAAACGGGAGACCAAGATCGCAAGCGCGGTACTGTCACCGAC 2040  
 Qy CGGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGACCAAGACCGAGCTGCGAG 2100  
 Db CGGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGACCAAGACCGAGCTGCGAG 2100  
 Qy GCGATCCAGCTGGCGCTGACGAGACGAGCGGAGCGAGGAGTGAACATCGTGACCGAGCGCAG 2160  
 Db GCGATCCAGCTGGCGCTGACGAGACGAGCGGAGCGAGGAGTGAACATCGTGACCGAGCGCAG 2160  
 Qy TACGCGCTGGCAGCATCATCAGCGCCAGCCGACAGAGAGCGAGCGAGCTGGTGAACCGAG 2220  
 Db TACGCGCTGGCAGCATCATCAGCGCCAGCCGACAGAGAGCGAGCGAGCTGGTGAACCGAG 2220  
 Qy ATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGTGGTCCCGCCACAG 2280  
 Db ATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGTGGTCCCGCCACAG 2280  
 Qy GGCAATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGCATCCGCAAGGTGCTG 2340  
 Db GGCAATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGCATCCGCAAGGTGCTG 2340  
 Qy TTCTCTGAGCGCATCGATGGCGGATCGTGATCTACCAAGTACATGAGCAGCCTGTACGCTG 2400  
 Db TTCTCTGAGCGCATCGATGGCGGATCGTGATCTACCAAGTACATGAGCAGCCTGTACGCTG 2400  
 Qy GGCAAGCGCGGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
 Db GGCAAGCGCGGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457

## RESULT 2

ADM73766

ID ADM73766 standard; DNA; 2457 BP.

XX AC ADM73766;

XX DT 03-JUN-2004 (first entry)

XX DE HIV-1 polynucleotide #9.

XX KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;

XX KW HIV type C protein; immunostimulant.

XX OS Human immunodeficiency virus 1.

XX PN US2003223961-A1.

XX PD 04-DEC-2003.

XX PF 05-JUL-2001; 2001US-00899575.

XX PR 05-JUL-2000; 2000US-00610313.

XX PA (MEGE/) MEGEDE J Z.

XX PA (BARN/) BARNETT S W.

XX PA (ENGE/) ENGELBRECHT S.

XX PA (RENS/) RENSBERG E J V.

XX PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;

XX PR WPI; 2004-060515/06.

XX PT New expression cassette comprising a polynucleotide sequence encoding an  
 PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
 PT immunization, generating of packaging cell lines or in producing HIV Type  
 PT C proteins.

XX PS Claim 1; SEQ ID NO 32; 160pp; English.

XX CC The invention relates to an expression cassette comprising a  
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention

CC also relates to a recombinant expression system for use in a host cell  
 CC comprising an expression cassette, where the polynucleotide sequence  
 CC further comprises control elements capable of driving expression in the  
 CC selected host cell, a cell comprising an expression cassette where the  
 CC polynucleotide sequence further comprises control elements compatible  
 CC with the expression in the cell and a composition for generating an  
 CC immunological response, comprising an expression cassette. The expression  
 CC cassette and the methods of the invention are useful in eliciting an  
 CC immune response, in DNA immunisation, in generation of packaging cell  
 CC lines and in producing HIV Type C proteins. This sequence represents an  
 CC HIV-1 polynucleotide of the invention.

SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 100.0%; Score 2457; DB 12; Length 2457;

Best Local Similarity 100.0%; Pred. No. 9.7e-296;

Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCACCATGGCGGAGGCGCATGAGCAGGCCACCGCCAAACATCCTGATGCAG 60  
 Db 1 GTCGACGCCACCATGGCGGAGGCGCATGAGCAGGCCACCGCCAAACATCCTGATGCAG 60  
 Qy 61 GCGAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGGCAAGAGGGC 120  
 Db 61 GCGAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGGCAAGAGGGC 120  
 Qy 121 CACATCCCGCGCAACTGCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
 Db 121 CACATCCCGCGCAACTGCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
 Qy 181 GGCACACAGATGAAGGACTGCGACCGAGCGCGCCAACTTCTTCGCGAGGACCTGGCC 240  
 Db 181 GGCACACAGATGAAGGACTGCGACCGAGCGCGCCAACTTCTTCGCGAGGACCTGGCC 240  
 Qy 241 TTCCCGCAGGCAAGGCGCGGAGTTCCCGAGCGAGCAGAACCGCGCCAAACAGCCCCACC 300  
 Db 241 TTCCCGCAGGCAAGGCGCGGAGTTCCCGAGCGAGCAGAACCGCGCCAAACAGCCCCACC 300  
 Qy 301 AGCGCGAGCTGCAGGTGCGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360  
 Db 301 AGCGCGAGCTGCAGGTGCGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360  
 Qy 361 GGCAACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGCATCAAGTG 420  
 Db 361 GGCAACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGCATCAAGTG 420  
 Qy 421 GCGGCGCAGATCAAGGAGGCGCTGCTGACACCGCGCGCGACGACACGCTGCTGGAGGAG 480  
 Db 421 GCGGCGCAGATCAAGGAGGCGCTGCTGACACCGCGCGCGACGACACGCTGCTGGAGGAG 480  
 Qy 481 ATGAGCCTGCGCGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG 540  
 Db 481 ATGAGCCTGCGCGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG 540  
 Qy 541 GTGCGCCAGTACGACACAGATCTGATCGAGATCTGCGGCAAGAGAGGCATCGGCACCGTG 600  
 Db 541 GTGCGCCAGTACGACACAGATCTGATCGAGATCTGCGGCAAGAGAGGCATCGGCACCGTG 600  
 Qy 601 CTGATCGGCGCCACCGCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC 660  
 Db 601 CTGATCGGCGCCACCGCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC 660  
 Qy 661 ACCCTGAACTTCCCGCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATG 720  
 Db 661 ACCCTGAACTTCCCGCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATG 720  
 Qy 721 GAGCGGCCCAAGGTGAAGCAGAGTGGCCCCCTGACCGAGGAGAGATCAAGAGGCCCTGACCGCC 780  
 Db 721 GAGCGGCCCAAGGTGAAGCAGAGTGGCCCCCTGACCGAGGAGAGATCAAGAGGCCCTGACCGCC 780  
 Qy 781 ATCTCGAGGAGATGGAGAGAGGAGGCAAGATCAACCAAGATCGGCCCCCGAGAACCCCTAC 840  
 Db 781 ATCTCGAGGAGATGGAGAGAGGAGGCAAGATCAACCAAGATCGGCCCCCGAGAACCCCTAC 840

QY 841 AACACCCCGCTGTTCCGCATCAAGAAAGAGACACCAAGTGGCGCAAGCTGTGGAC 900  
Db 841 AACACCCCGCTGTTCCGCATCAAGAAAGAGACACCAAGTGGCGCAAGCTGTGGAC 900  
QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCAC 960  
QY 961 CCCGCGGCTGAAGAAGAGAGCTGACCGTGTGGAGTGGGCGAGCGCTACTTC 1020  
Db 961 CCCGCGGCTGAAGAAGAGAGCGTGCACCGTGTGGAGTGGGCGAGCGCTACTTC 1020  
QY 1021 AGCGTGCCCTGGAGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
Db 1021 AGCGTGCCCTGGAGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
QY 1081 AACGAGACCCCGGATCCGCTACAGTAACAAGTGTGCTGCCAGGGCTGGAAGGCGAGC 1140  
Db 1081 AACGAGACCCCGGATCCGCTACAGTACAGTGTGCTGCCAGGGCTGGAAGGCGAGC 1140  
QY 1141 CCCAGCATCTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCGAACC 1200  
Db 1141 CCCAGCATCTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCGAACC 1200  
QY 1201 GAGATCGTGATCTACAGGCCCCCTGTAGTGGGCGAGCGCTGGAGATCGGCCAGCAC 1260  
Db 1201 GAGATCGTGATCTACAGGCCCCCTGTAGTGGGCGAGCGCTGGAGATCGGCCAGCAC 1260  
QY 1261 CGCGCAAGATCGAGGAGTGGCAAGCACCTGTGCGTGGGGCTTCCACACCCCGGAC 1320  
Db 1261 CGCGCAAGATCGAGGAGTGGCAAGCACCTGTGCGTGGGGCTTCCACACCCCGGAC 1320  
QY 1321 AAGAAGCACAGAGAGCCCTTCTGCCATCGAGCTGCACCCCGACAAAGTGAGCC 1380  
Db 1321 AAGAAGCACAGAGAGCCCTTCTGCCATCGAGCTGCACCCCGACAAAGTGAGCC 1380  
QY 1381 GTGCAGCCCATCGAGCTGCCGAGAGGAGCTGCAGCGTGAACGACATCCAGAGCTG 1440  
Db 1381 GTGCAGCCCATCGAGCTGCCGAGAGGAGCTGCAGCGTGAACGACATCCAGAGCTG 1440  
QY 1441 GTGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGC 1500  
Db 1441 GTGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGC 1500  
QY 1501 AAGCTGTGCGCGGCGCCAGGCCCTGACCGACATCTGTGCCCTGACCGAGGCGGAG 1560  
Db 1501 AAGCTGTGCGCGGCGCCAGGCCCTGACCGACATCTGTGCCCTGACCGAGGCGGAG 1560  
QY 1561 CTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGAGCCCGTGCACGGCGTGTACTACGAC 1620  
Db 1561 CTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGAGCCCGTGCACGGCGTGTACTACGAC 1620  
QY 1621 CCAGCAAGGACCTGTGGCGCAGATCCAGAGCAGGGCCAGCAGCTGGAACCTACCAG 1680  
Db 1621 CCAGCAAGGACCTGTGGCGCAGATCCAGAGCAGGGCCAGCAGCTGGAACCTACCAG 1680  
QY 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGCGACCGCC 1740  
Db 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGCGACCGCC 1740  
QY 1741 CACACCAAGCGTGAAGCGTGAACCGAGCCGTCAGAGATCGCCATGGAGAGCATC 1800  
Db 1741 CACACCAAGCGTGAAGCGTGAACCGAGCCGTCAGAGATCGCCATGGAGAGCATC 1800  
QY 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGGGAGACC 1860  
QY 1861 TGGTGAACGACTACTGGCAGGCCACTTGGATCCCGAGTGGGAGTTTGTGAACACCCCC 1920  
Db 1861 TGGTGAACGACTACTGGCAGGCCACTTGGATCCCGAGTGGGAGTTTGTGAACACCCCC 1920

QY 1921 CCCCTGTGAAGCTGTGTACAGCTGGAGAAAGAGCCCATCATCGGCGCGAGACCTTC 1980  
Db 1921 CCCCTGTGTGAAGCTGTGTACAGCTGGAGAAAGAGCCCATCATCGGCGCGAGACCTTC 1980  
QY 1981 TACGTGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGCTACGTGACCGAC 2040  
Db 1981 TACGTGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGCTACGTGACCGAC 2040  
QY 2041 CGGGCGCGGCAAGATCGTGAAGCTGACGAGACCAACCAACCAAGAGACCGAGCTGCAG 2100  
Db 2041 CGGGCGCGGCAAGATCGTGAAGCTGACGAGACCAACCAACCAAGAGACCGAGCTGCAG 2100  
QY 2101 GCCATCCAGCTGGCCCTGCAGACAGCGGCGAGCGAGTGAACATCGTGACCGACAGCCAG 2160  
Db 2101 GCCATCCAGCTGGCCCTGCAGACAGCGGCGAGCGAGTGAACATCGTGACCGACAGCCAG 2160  
QY 2161 TACGCTGTGGCATCATCCAGGCCAGCCCAAGAGCGAGCGAGCTGGTGAACCCAG 2220  
Db 2161 TACGCTGTGGCATCATCCAGGCCAGCCCAAGAGCGAGCGAGCTGGTGAACCCAG 2220  
QY 2221 ATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGCCCGCCACCAAG 2280  
Db 2221 ATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGCCCGCCACCAAG 2280  
QY 2281 GGCAATCGGCGCAACGAGCAGATCGACAAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340  
Db 2281 GGCAATCGGCGCAACGAGCAGATCGACAAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG 2340  
QY 2341 TTCTGTGACGCGATCGATGGCGGCATCGTATCTACAGTACATGAGCAGCTGTACGCTG 2400  
Db 2341 TTCTGTGACGCGATCGATGGCGGCATCGTATCTACAGTACATGAGCAGCTGTACGCTG 2400  
QY 2401 GGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
Db 2401 GGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457

RESULT 3

ACA03546  
ID ACA03546 standard; DNA; 2445 BP.

AC ACA03546;

XX AC

DT 22-MAY-2003 (first entry)

XX DE Synthetic DNA encoding immunogenic HIV peptide #29.

XX DE Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.

XX OS Synthetic.

XX PN WO2003004657-A1.

XX PD 16-JAN-2003.

XX PF 05-JUL-2002; 2002WO-US021421.

XX PR 05-JUL-2001; 2001US-0303192P.

XX PR 31-AUG-2001; 2001US-0316860P.

XX PR 16-JAN-2002; 2002US-0349728P.

XX PR 16-JAN-2002; 2002US-0349793P.

XX PR 16-JAN-2002; 2002US-0349871P.

XX PA (CHIR ) CHIRON CORP.

XX PI Zur Megede J, Barnett SW, Lian Y;

XX DR WPI; 2003-221602/21.

XX PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C polypeptides, useful as immunogenic compositions or vaccines for

generating humoral or cellular immune responses against HIV in a subject, especially humans.

Example 1; Fig 34; 262pp; English.

The invention describes a synthetic polynucleotide encoding 2 or more immunogenic HIV polypeptides, where at least 2 of the polypeptides are derived from different HIV subtypes. The polynucleotide is useful for immunisation, generation of packaging cell lines, or production of HIV polypeptides. The polynucleotide and its encoded proteins are useful as immunogenic compositions or vaccines for generating humoral or cellular immune responses against HIV in a subject, or for inducing neutralising antibodies against HIV. The gene delivery vector comprising the polynucleotide is also useful for DNA immunisation of, or for generating an immune response (e.g. a humoral or cellular immune response) in, a subject such as a mammal, particularly a human. This sequence encodes a human immunodeficiency virus immunogenic peptide

Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 99.4%; Score 2443.4; DB 8; Length 2445;

Best Local Similarity 100.0%; Pred. No. 4.7e-294;

Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	7	GCCACCATGGCCGAGGCCATGAGCCAGGCCACACAGCGCCAACTCTGATGCGCGCAGC	66
Db	1	GCCACCATGGCCGAGGCCATGAGCCAGGCCACACAGCGCCAACTCTGATGCGCGCAGC	60
Qy	67	AACCTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCACTCGCGCAGAGGGGCCACATC	126
Db	61	AACCTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCACTCGCGCAGAGGGGCCACATC	120
Qy	127	GCCGCGCAACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGGCAAGGGGGCCAC	186
Db	121	GCCGCGCAACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGGCAAGGGGGCCAC	180
Qy	187	CAGATGAAGGACTGCACCGAGCGCCAGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCC	246
Db	181	CAGATGAAGGACTGCACCGAGCGCCAGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCC	240
Qy	247	CAGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCAAACAGCCCCCAGCGCG	306
Db	241	CAGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCAAACAGCCCCCAGCGCG	300
Qy	307	GAGCTGCAGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCAC	366
Db	301	GAGCTGCAGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCAC	360
Qy	367	CTGAACTTCCCCCAGATCACCTTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC	426
Db	361	CTGAACTTCCCCCAGATCACCTTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC	420
Qy	427	CAGATCAAGGAGGCCCTGTGGACACCGCGCGCCGACGACACCGTGTGGAGGAGATGAGC	486
Db	421	CAGATCAAGGAGGCCCTGTGGACACCGCGCGCCGACGACACCGTGTGGAGGAGATGAGC	480
Qy	487	CTGCGCGCAGTGGAGGCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGGC	546
Db	481	CTGCGCGCAGTGGAGGCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGGC	540
Qy	547	CAGTACGACCATGCTGTGATCGAGATCTGCGCGCAAGAGGCCATCGCGCACCGTCTGTATC	606
Db	541	CAGTACGACCATGCTGTGATCGAGATCTGCGCGCAAGAGGCCATCGCGCACCGTCTGTATC	600
Qy	607	GGCCCCACCCCGTGAAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGACCCCTG	666
Db	601	GGCCCCACCCCGTGAAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGACCCCTG	660
Qy	667	AACCTTCCCATCAGCCCCATCGACCGTGGCCCTGAGCTGAGCCCGGCGATGGAGCGC	726
Db	661	AACCTTCCCATCAGCCCCATCGACCGTGGCCCTGAGCTGAGCCCGGCGATGGAGCGC	720
Qy	727	CCCAAGGTGAAGCAGTGGCCCTTGACCGGAGGAGATCAAGGGCCCTTGACCGCCATCTGC	786

Db	721	CCCAAGGTGAAGCAGTGGCCCTTGACCGGAGGAGATCAAGGGCCCTGACCGCCATCTGC	780
Qy	787	GAGGAGATGAGAGGAGGGCCAGATCACAAGATCGGCCCGGAGNACCCCTACAAACACC	846
Db	781	GAGGAGATGAGAGGAGGGCCAGATCACAAGATCGGCCCGGAGNACCCCTACAAACACC	840
Qy	847	CCCGTGTTCGCCATCAAGAAGAAGACAGACCAAGTGGCGCAAGCTGGTGGATTTCCGC	906
Db	841	CCCGTGTTCGCCATCAAGAAGAAGACAGACCAAGTGGCGCAAGCTGGTGGATTTCCGC	900
Qy	907	GAGCTGAACCAAGGCCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCACCCCGCC	966
Db	901	GAGCTGAACCAAGGCCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCACCCCGCC	960
Qy	967	GGCTGGAAGAGAGAGAGCGGTGACCGTGTGACCTGGCGGACGCGCTACTTCAGCGTG	1026
Db	961	GGCTGGAAGAGAGAGAGCGGTGACCGTGTGACCTGGCGGACGCGCTACTTCAGCGTG	1020
Qy	1027	CCCTGGACGAGGACTTCCGCAAGTACACCGCCCTTCAACCATCCCGCAGCATCAACAAAGAG	1086
Db	1021	CCCTGGACGAGGACTTCCGCAAGTACACCGCCCTTCAACCATCCCGCAGCATCAACAAAGAG	1080
Qy	1087	ACCCCGGCGATTCGCTTACCAAGTACAAAGTGTGCTGCCAGGGCTGGAGGGCAGCCCCCAGC	1146
Db	1081	ACCCCGGCGATTCGCTTACCAAGTACAAAGTGTGCTGCCAGGGCTGGAGGGCAGCCCCCAGC	1140
Qy	1147	ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC	1206
Db	1141	ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC	1200
Qy	1207	GTGATCTTACCAAGGCCCCCTGTACGTGGGAGCAGCCTGTGAGATCGGCGACACCGCGCC	1266
Db	1201	GTGATCTTACCAAGGCCCCCTGTACGTGGGAGCAGCCTGTGAGATCGGCGACACCGCGCC	1260
Qy	1267	AAGATCGAGGAGTGTGCGCAAGCACCTGCTGCGTGGGGCTTTCACACCCCCCGCAAGAAG	1326
Db	1261	AAGATCGAGGAGTGTGCGCAAGCACCTGCTGCGTGGGGCTTTCACACCCCCCGCAAGAAG	1320
Qy	1327	CACCGAGAGGAGCCCTTCTTCCGCGATCGAGCTGACCCCGACCAAGTGCAGCTGCAG	1386
Db	1321	CACCGAGAGGAGCCCTTCTTCCGCGATCGAGCTGACCCCGACCAAGTGCAGCTGCAG	1380
Qy	1387	CCCATTCAGCTGCGCGGAGAGGAGCTGGAACCGTGAACGACATCCAGAACCTTGTGGGC	1446
Db	1381	CCCATTCAGCTGCGCGGAGAGGAGCTGGAACCGTGAACGACATCCAGAACCTTGTGGGC	1440
Qy	1447	AAGCTGAATCTGGCCAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG	1506
Db	1441	AAGCTGAATCTGGCCAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG	1500
Qy	1507	CTGGCGGCGCCAGGCGCTTACCGACATCGTGGCCCTGACCGAGGAGGCGGAGCTGGAG	1566
Db	1501	CTGGCGGCGCCAGGCGCTTACCGACATCGTGGCCCTGACCGAGGAGGCGGAGCTGGAG	1560
Qy	1567	CTGGCCGAGAACCCGCGAGATCTTGGCGGAGCCCGTGCACGCGCTGTACTACAGCCCAAGC	1626
Db	1561	CTGGCCGAGAACCCGCGAGATCTTGGCGGAGCCCGTGCACGCGCTGTACTACAGCCCAAGC	1620
Qy	1627	AAGSACCTGGTGGCCGAGATCCAGAAGCAGGGGCCACACAGTGGAGCTTACAGATCTTAC	1686
Db	1621	AAGSACCTGGTGGCCGAGATCCAGAAGCAGGGGCCACACAGTGGAGCTTACAGATCTTAC	1680
Qy	1687	CAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCGCAAGATGCGCACCGGCCACACC	1746
Db	1681	CAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTACGCGCAAGATGCGCACCGGCCACACC	1740
Qy	1747	AACGACGTGAAGCAGCTGACCGAGGCGCGTGCAGAAAGATCGCCATGGAGAGCATCTGTATC	1806
Db	1741	AACGACGTGAAGCAGCTGACCGAGGCGCGTGCAGAAAGATCGCCATGGAGAGCATCTGTATC	1800
Qy	1807	TGGGGCAAGACCCCAAGTTCCCGCTTCCCATTCAGAAAGGAGACCTGGGAGACCTGGTGG	1866



Db 1801 TGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGGAGACTGGTGG 1860  
 Qy 1867 ACCGACTACTGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAAACACCCCCCTG 1926  
 Db 1861 ACCGACTACTGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAAACACCCCCCTG 1920  
 Qy 1927 GTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTG 1986  
 Db 1921 GTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTG 1980  
 Qy 1987 GACGCGCGCGCAACCGGAGACCAAGATCGGCAAGGCGGCTACGTCAACGACCGGGC 2046  
 Db 1981 GACGCGCGCGCAACCGGAGACCAAGATCGGCAAGGCGGCTACGTCAACGACCGGGC 2040  
 Qy 2047 CGGAGAGATGCTGAGCTGACCGAGACCAAGACCAACCAAGAGACGAGCTGCAGGCCATC 2106  
 Db 2041 CGGAGAGATGCTGAGCTGACCGAGACCAACCAAGAGACGAGCTGCAGGCCATC 2100  
 Qy 2107 CAGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTGACGACAGCCAGTACGCC 2166  
 Db 2101 CAGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTGACGACAGCCAGTACGCC 2160  
 Qy 2167 CTGGGCATCATCCAGGCCAGCCGACAGAGCGAGCGAGCTGGTGAACCCAGATCATC 2226  
 Db 2161 CTGGGCATCATCCAGGCCAGCCGACAGAGCGAGCGAGCTGGTGAACCCAGATCATC 2220  
 Qy 2227 GAGGAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGGTCCCGCCACAGGGGCATC 2286  
 Db 2221 GAGGAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGGTCCCGCCACAGGGGCATC 2280  
 Qy 2287 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGGCATCCGCAAGTGTCTTCTG 2346  
 Db 2281 GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGGCATCCGCAAGTGTCTTCTG 2340  
 Qy 2347 GACGCGATCGATGGCGGCATCGTATCTACCAAGTACATGAGCAGCTGTACGTGGGCAGC 2406  
 Db 2341 GACGCGATCGATGGCGGCATCGTATCTACCAAGTACATGAGCAGCTGTACGTGGGCAGC 2400  
 Qy 2407 GCGGCGCTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
 Db 2401 GCGGCGCTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2445

RESULT 4  
 ADCL3264  
 ID ADCL3264 standard; DNA; 2445 BP.  
 AC ADCL3264;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE DNA of HIV construct p2Pol-opt-YMM\_C SEQ ID NO 43.  
 XX  
 KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; da.  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 PN WO2003004620-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021420.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA (UYST-) UNIV STELLENBOSCH.  
 XX  
 PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 XX

DR WPI; 2003-221593/21.  
 XX  
 PT New expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide, useful for immunization, or generating packaging cell lines.  
 PT  
 XX  
 PS Disclosure; Fig 40; 301pp; English.  
 XX  
 CC The invention relates to a novel expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel expression cassette can be used to treat HIV type C by gene therapy or used in the development of a vaccine. The gene delivery vector is administered intramuscularly, intramuscularly, intranasally, subcutaneously, intradermally, transdermally, intravenously, intrarectally, orally or intravenously. The expression cassette is useful for immunisation, generating packaging cell lines and producing HIV polypeptides. This polynucleotide sequence represents the DNA of an HIV Type C related sequence of the invention.  
 CC  
 XX  
 SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;  
 Query Match 99.4%; Score 2443.4; DB 10; Length 2445;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-294;  
 Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 7 GCCACCATGGCGGAGCCCATGAGCCAGCCACAGCGCCCAACATCTCTATGAGCGCAGC 66  
 Db 1 GCCACCATGGCGGAGCCCATGAGCCAGCCACAGCGCCCAACATCTCTATGAGCGCAGC 60  
 Qy 67 AACTTCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTCGGCAAGGCGCCACATC 126  
 Db 61 AACTTCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTCGGCAAGGCGCCACATC 120  
 Qy 127 GCGCGCAACTCGCGCGCCCGCCGCAAGAGGGCTGTCTGAAGTGTGGCAAGAGGGCCAC 186  
 Db 121 GCGCGCAACTCGCGCGCCCGCCGCAAGAGGGCTGTCTGAAGTGTGGCAAGAGGGCCAC 180  
 Qy 187 CAGATGAAGGACTGACCGAGCGCGCCGCAACATCTTCTCCGCGAGGACTGTGCTTCCC 246  
 Db 181 CAGATGAAGGACTGACCGAGCGCGCCGCAACATCTTCTCCGCGAGGACTGTGCTTCCC 240  
 Qy 247 CAGGCAAGGCGCGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCGCAGCGC 306  
 Db 241 CAGGCAAGGCGCGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCGCAGCGC 300  
 Qy 307 GAGCTGCAAGTGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCACC 366  
 Db 301 GAGCTGCAAGTGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCACC 360  
 Qy 367 CTGAATCTTCCCGCAGATCACCTGTGTGAGCGCGCCCTCTGTGTAGCATCAAGTGGCGCGC 426  
 Db 361 CTGAATCTTCCCGCAGATCACCTGTGTGAGCGCGCCCTCTGTGTAGCATCAAGTGGCGCGC 420  
 Qy 427 CAGATCAAGGAGCGCTCTGTGACACCGCGCGCAGACACCGTCTGTGAGAGATGAGC 486  
 Db 421 CAGATCAAGGAGCGCTCTGTGACACCGCGCGCAGACACCGTCTGTGAGAGATGAGC 480  
 Qy 487 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGTGGCGC 546  
 Db 481 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGTGGCGC 540  
 Qy 547 CAGTACGACCAAGATCTCTGATGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATC 606  
 Db 541 CAGTACGACCAAGATCTCTGATGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATC 600  
 Qy 607 G3CCCCACCCCTGGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGTGACCCCTG 666  
 Db 601 G3CCCCACCCCTGGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGTGACCCCTG 660  
 Qy 667 AACTTCCCATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATGACCGGC 726

Db 661 AACTTCCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGGAACGGC 720  
Qy 727 CCAAGGTGAAGCAGTGGCCCCCTGACGAGGAGAGATCAAGGGCCCTGACCGCCATCTGC 786  
Db 721 CCAAGGTGAAGCAGTGGCCCCCTGACGAGGAGAGATCAAGGGCCCTGACCGCCATCTGC 780  
Qy 787 GAGGAGATGGAAGAGAGGGGCAAGATCAACGAAGATCGGCCCGGAGAACCCCTCAACAACC 846  
Db 781 GAGGAGATGGAAGAGAGGGGCAAGATCAACGAAGATCGGCCCGGAGAACCCCTCAACAACC 840  
Qy 847 CCGGTGTTCCGCATCAAGAGAAGGAGCAGCACCAGTGGCGCAAGCTGGTGATCTTCGC 906  
Db 841 CCGGTGTTCCGCATCAAGAGAAGGAGCAGCACCAGTGGCGCAAGCTGGTGATCTTCGC 900  
Qy 907 GAGCTGAACAAAGCGCACCAGGACTTCTGGGAGGTGACAGTGGGCGATCCCCCACCAGCCGCC 966  
Db 901 GAGCTGAACAAAGCGCACCAGGACTTCTGGGAGGTGACAGTGGGCGATCCCCCACCAGCCGCC 960  
Qy 967 GGCCTGAAGAGAAGAGAGCGTGACCGTGTGACAGTGGGCGAGCGCTACTTTCAGCGTG 1026  
Db 961 GGCCTGAAGAGAAGAGAGCGTGACCGTGTGACAGTGGGCGAGCGCTACTTTCAGCGTG 1020  
Qy 1027 CCCCTGAGCAGGAGACTTCCGCAAGTACACCCGCTTCAACATCCCGAGCATCAACAACGAG 1086  
Db 1021 CCCCTGAGCAGGAGACTTCCGCAAGTACACCCGCTTCAACATCCCGAGCATCAACAACGAG 1080  
Qy 1087 ACCCCCGGCATCCGCTACAGTACAAGGTGCTGCCCGAGGGCTGGAAGGGCAGCCCGAGC 1146  
Db 1081 ACCCCCGGCATCCGCTACAGTACAAGGTGCTGCCCGAGGGCTGGAAGGGCAGCCCGAGC 1140  
Qy 1147 ATCTTCAGAGCAGATGACCAAGATCTGAGAGCCCTTTCGCGCGCGGCAACCCCGAGATC 1206  
Db 1141 ATCTTCAGAGCAGATGACCAAGATCTGAGAGCCCTTTCGCGCGCGGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACAGGCGCCCTCTGAGTGGGAGCGACCTGGAGATCGGCGACGCGCGCC 1266  
Db 1201 GTGATCTACAGGCGCCCTCTGAGTGGGAGCGACCTGGAGATCGGCGACGCGCGCC 1260  
Qy 1267 AAGATCGAGGAGCTGCGCAAGCAGCTGCTGGCGTGGGGCTTCAACACCCCGCAAGAAG 1326  
Db 1261 AAGATCGAGGAGCTGCGCAAGCAGCTGCTGGCGTGGGGCTTCAACACCCCGCAAGAAG 1320  
Qy 1327 CACCAAGAGGAGCCCTTCTGCGCATGAGCTGCAACCCCGCAAGAAGTGGACCGTGAG 1386  
Db 1321 CACCAAGAGGAGCCCTTCTGCGCATGAGCTGCAACCCCGCAAGAAGTGGACCGTGAG 1380  
Qy 1387 CCCATCGAGCTGCGGAGAGAGCTGAGACCTGGAACGACATCCAGAGCTGGTGGC 1446  
Db 1381 CCCATCGAGCTGCGGAGAGAGCTGAGACCTGGAACGACATCCAGAGCTGGTGGC 1440  
Qy 1447 AAGCTGAACCTGGGCGAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506  
Db 1441 AAGCTGAACCTGGGCGAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500  
Qy 1507 CTGCGCGCGCCAAAGGCCCTGACCGACATGTGCCCCCTGACCCGAGGAGCGGAGCTGGAG 1566  
Db 1501 CTGCGCGCGCCAAAGGCCCTGACCGACATGTGCCCCCTGACCCGAGGAGCGGAGCTGGAG 1560  
Qy 1567 CTGCGCGAGAACCCGAGATCTCTGCGGAGCCCTGACCGCGTGTACTAGACCCCGAGC 1626  
Db 1561 CTGCGCGAGAACCCGAGATCTCTGCGGAGCCCTGACCGCGTGTACTAGACCCCGAGC 1620  
Qy 1627 AAGGACCTGTGGCGGAGATCCAGAAGCAGGGCCACGACAGTGGACCTACAGATCTAC 1686  
Db 1621 AAGGACCTGTGGCGGAGATCCAGAAGCAGGGCCACGACAGTGGACCTACAGATCTAC 1680  
Qy 1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGGCCACAC 1746  
Db 1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGGCCACAC 1740  
Qy 1747 AACGACGTGAAGCAGCTGACCGAGGCGGTGAGAGATCGCCATGGAAGCATCGTGATC 1806  
Db 1741 AACGACGTGAAGCAGCTGACCGAGGCGGTGAGAGATCGCCATGGAAGCATCGTGATC 1800

Qy 1807 TGGGCAAGACCCCACAGTTCCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGG 1866  
Db 1801 TGGGCAAGACCCCACAGTTCCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGG 1860  
Qy 1867 ACCGACTACTGGGAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCCCTG 1926  
Db 1861 ACCGACTACTGGGAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCCCTG 1920  
Qy 1927 GTGAAGCTGTGGTACCAAGTGGAGAGGCCCATCATCGGCGCGAGACCTTCTACGTG 1986  
Db 1921 GTGAAGCTGTGGTACCAAGTGGAGAGGCCCATCATCGGCGCGAGACCTTCTACGTG 1980  
Qy 1987 GACGCGCCCGCCAAACCGCGAGACCAAGATCGGCAAGGCCGCTACTAGTGAACCGAGCCGGGC 2046  
Db 1981 GACGCGCCCGCCAAACCGCGAGACCAAGATCGGCAAGGCCGCTACTAGTGAACCGAGCCGGGC 2040  
Qy 2047 CGGCAAGAATCTGTGAGCCCTGACCGAGACCAACCAACAGAGACCGAGCTGAGGCCATC 2106  
Db 2041 CGGCAAGAATCTGTGAGCCCTGACCGAGACCAACCAACAGAGACCGAGCTGAGGCCATC 2100  
Qy 2107 CAGCTGGCCCTGCAAGGACAGCGGAGCGAGGTGAACATCGTGACCGAGCAGCAGTACGCC 2166  
Db 2101 CAGCTGGCCCTGCAAGGACAGCGGAGCGAGGTGAACATCGTGACCGAGCAGCAGTACGCC 2160  
Qy 2167 CTGGGCAATCATCCAGGCCCCAGCCGACAGAGCGAGGCGAGCTGGTGAACCCAGATCATC 2226  
Db 2161 CTGGGCAATCATCCAGGCCCCAGCCGACAGAGCGAGGCGAGCTGGTGAACCCAGATCATC 2220  
Qy 2227 GAGCAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGGTGCCCGGCCCAAGAGGCCATC 2286  
Db 2221 GAGCAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGGTGCCCGGCCCAAGAGGCCATC 2280  
Qy 2287 GGGGCAACAGCAGCAGATCGCAAGCTGTGAGCAAGGGGCATCCGCAAGGTGCTGTCTCTG 2346  
Db 2281 GGGGCAACAGCAGCAGATCGCAAGCTGTGAGCAAGGGGCATCCGCAAGGTGCTGTCTCTG 2340  
Qy 2347 GAGCGCATCGATGCGGCGCATCGTGTATACCAAGTACATGACGACACCTGTAGTGGGAGC 2406  
Db 2341 GAGCGCATCGATGCGGCGCATCGTGTATACCAAGTACATGACGACACCTGTAGTGGGAGC 2400  
Qy 2407 GGGGCGCCCTAGGATCGATTAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 2401 GGGGCGCCCTAGGATCGATTAAGCTTCCCGGGCTAGCACCGGT 2445

## RESULT 5

ADC13230  
ID ADC13230 standard; DNA; 3930 BP.

XX AC ADC13230;

XX DT 18-DEC-2003 (first entry)

XX DNA of HIV construct GagCompIPolmut\_C SEQ ID NO 9.

XX DE expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
XX KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX OS Human immunodeficiency virus.

XX PN WO2003004620-A2.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.

XX 31-AUG-2001; 2001US-0316860P.

XX 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

PA (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 PI WPI; 2003-221593/21.  
 XX  
 PT New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.  
 XX  
 PS Disclosure; Fig 6; 301pp; English.  
 PS  
 XX  
 CC The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intramuscularly, intranasally,  
 CC subcutaneously, intradermally, transdermally, intravaginally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.  
 XX  
 SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.4; DB 10; Length 3930;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-293;  
 Matches 2437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 14 TGGCCGAGCCATGAGCCAGGCGCCAGCCAGCCCAACATCTGATGAGCGCAGCACTTCA 73  
 Db 1487 TGGCCGAGCCATGAGCCAGGCGCCAGCCAGCCCAACATCTGATGAGCGCAGCACTTCA 1546

QY 74 AGGGCCCCAAGGCGATCATCAAGTCTTCACTGCGGCAAGAGAGGGCCACATCGCCGCA 133  
 Db 1547 AGGGCCCCAAGGCGATCATCAAGTCTTCACTGCGGCAAGAGAGGGCCACATCGCCGCA 1606

QY 134 ACTGCGGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGGCAAGAGGGCCACAGATGA 193  
 Db 1607 ACTGCGGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGGCAAGAGGGCCACAGATGA 1666

QY 194 AGGACTGCAACCGAGCGCAGGCGCAACTTCTCCGAGGAGCTGCGCTTCCCGCAGGCA 253  
 Db 1667 AGGACTGCAACCGAGCGCAGGCGCAACTTCTCCGAGGAGCTGCGCTTCCCGCAGGCA 1726

QY 254 AGGCGCGGAGTTCCCGAGCGAGAGAACCGCGCCAAAGCCCAAGCGCGAGCTGC 313  
 Db 1727 AGGCGCGGAGTTCCCGAGCGAGAGAACCGCGCCAAAGCCCAAGCGCGAGCTGC 1786

QY 314 AGGTGCGGCGGCAACACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGGAACT 373  
 Db 1787 AGGTGCGGCGGCAACACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGGAACT 1846

QY 374 TCCCCAGATCACTCTGTGGAGCGCCCTCTGAGGATCAAGGTGGCGGCGCAGATCA 433  
 Db 1847 TCCCCAGATCACTCTGTGGAGCGCCCTCTGAGGATCAAGGTGGCGGCGCAGATCA 1906

QY 434 AGGAGGCGCTGTGGACACCGCGCGCGAGCACACCGTGTGGAGGAGATGAGCTGCGCG 493  
 Db 1907 AGGAGGCGCTGTGGACACCGCGCGCGAGCACACCGTGTGGAGGAGATGAGCTGCGCG 1966

QY 494 GCAAGTGGAGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGGCGCAGTAGG 553  
 Db 1967 GCAAGTGGAGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGGCGCAGTAGG 2026

QY 554 ACCAGATCTGATCGAGATCTCGGCAAGAGGGCCATCGGCACTGCTGATCGGCCCA 613  
 Db 2027 ACCAGATCTGATCGAGATCTCGGCAAGAGGGCCATCGGCACTGCTGATCGGCCCA 2086

QY 614 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACTTCACTTCC 673  
 Db 2087 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACTTCACTTCC 2146

QY 674 CCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGATGAGCGCCCCAAGG 733  
 Db 2147 CCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGATGAGCGCCCCAAGG 2206

QY 734 TGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCGCTTGAACGCCATCTGCGAGGAGA 793  
 Db 2207 TGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCGCTTGAACGCCATCTGCGAGGAGA 2266

QY 794 TGGAGAGGAGGGCAAGATCAACAGATCGCGCCCGAGAACCCCTTACAACACCCCGTGT 853  
 Db 2267 TGGAGAGGAGGGCAAGATCAACAGATCGCGCCCGAGAACCCCTTACAACACCCCGTGT 2326

QY 854 TCGCCATCAAGNAGAGGAGGACCAAGTGGCGGAGCTGGTGGAGCTTCGCGAGCTGA 913  
 Db 2327 TCGCCATCAAGNAGAGGAGGACCAAGTGGCGGAGCTGGTGGAGCTTCGCGAGCTGA 2386

QY 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGACATCGGGGATCCCCACCCCGCGGCTGA 973  
 Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGACATCGGGGATCCCCACCCCGCGGCTGA 2446

QY 974 AGAAGAAAGAGCGGTGACCGTGTGAGAGTGGGCGACGCTTACTTCAAGCTGCGCCCTGG 1033  
 Db 2447 AGAAGAAAGAGCGGTGACCGTGTGAGAGTGGGCGACGCTTACTTCAAGCTGCGCCCTGG 2506

QY 1034 ACGAGGACTTCGCGAAGTACACCGCTTCAACATCCCGAGCATCAACACGAGACCCCG 1093  
 Db 2507 ACGAGGACTTCGCGAAGTACACCGCTTCAACATCCCGAGCATCAACACGAGACCCCG 2566

QY 1094 GCATCCGCTACAGTACAAACGCTGCTGCCCGAGGCTGGAAGGCGAGCCAGCATCTTCC 1153  
 Db 2567 GCATCCGCTACAGTACAAACGCTGCTGCCCGAGGCTGGAAGGCGAGCCAGCATCTTCC 2626

QY 1154 AGAGCAGCATGACCAAGATCTTGAGGCTTCCCGCGCCCGCAACCCCGAGATCTGATCT 1213  
 Db 2627 AGAGCAGCATGACCAAGATCTTGAGGCTTCCCGCGCCCGCAACCCCGAGATCTGATCT 2686

QY 1214 ACCAGGCCCCCTGTACGTGGGCGAGCATCTGGAGTGGCGGAGCGGCGGAGATCG 1273  
 Db 2687 ACCAGGCCCCCTGTACGTGGGCGAGCATCTGGAGTGGCGGAGCGGCGGAGATCG 2746

QY 1274 AGGAGCTCGCAAGCACCTGCTGGCGTGGGCTTCAACACCCCGCAAGAGCACCA 1333  
 Db 2747 AGGAGCTCGCAAGCACCTGCTGGCGTGGGCTTCAACACCCCGCAAGAGCACCA 2806

QY 1334 AGGAGCCCCCTTCTGCGCATCGAGCTGCAACCGCGAACAGTGGACCGTGAGCCCATCG 1393  
 Db 2807 AGGAGCCCCCTTCTGCGCATCGAGCTGCAACCGCGAACAGTGGACCGTGAGCCCATCG 2866

QY 1394 AGCTGCCGAGNAGGAGAGCTGGACCGTGAACGACATCCAGAACTGCTGGCAAGCTGA 1453  
 Db 2867 AGCTGCCGAGNAGGAGAGCTGGACCGTGAACGACATCCAGAACTGCTGGCAAGCTGA 2926

QY 1454 ACTGGGCGAGCAGATCTTACCCCGGCTTCAAGGCTGGCGGAGCTGCAAGCTGCTGCGG 1513  
 Db 2927 ACTGGGCGAGCAGATCTTACCCCGGCTTCAAGGCTGGCGGAGCTGCAAGCTGCTGCGG 2986

QY 1514 GCGCAAGGCGCTGACCGCATCTGTCCTTGAACCGAGGAGCGCGAGCTGAGCTGCGCG 1573  
 Db 2987 GCGCAAGGCGCTGACCGCATCTGTCCTTGAACCGAGGAGCGCGAGCTGAGCTGCGCG 3046

QY 1574 AGAACCGGAGATCTGCGCGAGCCCGTGCACGGGCTGTACTACGACCCAGCAAGGAGCC 1633  
 Db 3047 AGAACCGGAGATCTGCGCGAGCCCGTGCACGGGCTGTACTACGACCCAGCAAGGAGCC 3106

QY 1634 TGGTGGCGAGATCCAGAAAGAGGCGCACGACAGTGGACCTTACAGATCTTACAGGAGC 1693  
 Db 3107 TGGTGGCGAGATCCAGAAAGAGGCGCACGACAGTGGACCTTACAGATCTTACAGGAGC 3166

QY 1694 CTTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATCGGCAACCGCCCAACCAACGAG 1753  
 Db 3167 CTTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATCGGCAACCGCCCAACCAACGAG 3226

```
Qy 1754 TGAAGCAGCTGACCGAGGCGCTGCAGAGAGATCGCCATGAGAGCATCGTATCTGGGGCA 1813
Db 3227 TGAAGCAGCTGACCGAGGCGCTGCAGAGAGATCGCCATGAGAGCATCGTATCTGGGGCA 3286
Qy 1814 AGACCCCAAGCTTCGGCTGCCATCCAGAGAGAGACTGGGAGACCTGGTGGACCGACT 1873
Db 3287 AGACCCCAAGCTTCGGCTGCCATCCAGAGAGAGACTGGGAGACCTGGTGGACCGACT 3346
Qy 1874 ACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTAACACACCCCGCTGGTGAAGC 1933
Db 3347 ACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTAACACACCCCGCTGGTGAAGC 3406
Qy 1934 TGTGGTACCAGCTGGAGAGAGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGCG 1993
Db 3407 TGTGGTACCAGCTGGAGAGAGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGCG 3466
Qy 1994 CCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGAGCGGGCGGCGAGA 2053
Db 3467 CCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGAGCGGGCGGCGAGA 3526
Qy 2054 AGATCGTGAGCTGACCGAGACCAACCAAGAGAGACCGAGCTGACGCGCATCCAGCTGG 2113
Db 3527 AGATCGTGAGCTGACCGAGACCAACCAAGAGAGACCGAGCTGACGCGCATCCAGCTGG 3586
Qy 2114 CCCTGCGAGGACGCGGAGCGAGGTGAACATCGTGACCGAGCGAGTACGCCCTGGGCA 2173
Db 3587 CCCTGCGAGGACGCGGAGCGAGGTGAACATCGTGACCGAGCGAGTACGCCCTGGGCA 3646
Qy 2174 TCATCGAGGCGCGAGCGAGAGCGAGCGAGCTGGTGAACAGATCATCGAGCAGC 2233
Db 3647 TCATCGAGGCGCGAGCGAGAGCGAGCGAGCTGGTGAACAGATCATCGAGCAGC 3706
Qy 2234 TGATCAAGAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATCGCGGCA 2293
Db 3707 TGATCAAGAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATCGCGGCA 3766
Qy 2294 ACGAGCAGATCGACAAGCTGTGTAGCAAGGCGATCCCGCAAGGTCTGTCTCTGACGCGCA 2353
Db 3767 ACGAGCAGATCGACAAGCTGTGTAGCAAGGCGATCCCGCAAGGTCTGTCTCTGACGCGCA 3826
Qy 2354 TCGATGCGGCGCATCGTATCTACCACTACATGAGACGACCTGTAGCTGGGCGAGCGCGCC 2413
Db 3827 TCGATGCGGCGCATCGTATCTACCACTACATGAGACGACCTGTAGCTGGGCGAGCGCGCC 3886
Qy 2414 CTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db 3887 CTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 6
ABL39960
ID ABL39960 standard; DNA; 2463 BP.
XX ABL39960;
AC
XX
XX
DT 15-MAY-2002 (first entry)
XX
XX
DE Synthetic construct PR975YM SEQ ID NO:31.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
XX Human immunodeficiency virus; type C.
OS Synthetic.
XX
XX WO200204493-A2.
XX
XX 17-JAN-2002.
XX
XX 05-JUL-2001; 2001WO-US021241.
PF
XX
XX 05-JUL-2000; 2000US-00610313.
PR
```

XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2002-154920/20.

XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in  
PT applications including DNA immunization or generation of packaging cell  
PT lines, particularly in gene therapy.

XX Claim 1; Fig 9; 233pp; English.

XX The present invention describes expression cassettes comprising a  
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV  
CC type C polypeptides. The expression cassettes comprise any of the HIV  
CC type C sequences encoding gag, pol, vif, vpr, tat, rev, vpu, env or nef  
CC (i). (i) have immunostimulant activity and can be used in gene therapy.  
CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC Type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABL06204 to ABL06215 represent sequences used in the exemplification of  
CC the present invention

XX SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.2; DB 6; Length 2463;  
Best Local Similarity 99.6%; Pred. No. 3.6e-293;  
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy 1 GTGAGCGCCACCATGGCGGAGGCATGAGCGAGCCAGCGAGCGCAACATCCTGATGCGAG 60
Db 1 GTGAGCGCCACCATGGCGGAGGCATGAGCGAGCCAGCGAGCGCAACATCCTGATGCGAG 60
Qy 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTCGCGCAAGAGAGGC 120
Db 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTCGCGCAAGAGAGGC 120
Qy 121 CATATGCGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 121 CATATGCGCGCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Qy 181 GGGCACCAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 GGGCACCAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Qy 241 TTCCCCCAGGCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 241 TTCCCCCAGGCGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Qy 301 AGCGCGAGCTGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 301 AGCGCGAGCTGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Qy 361 GGCACCTGCAACTTCCCGCGAGATCACTCTGCGAGCGCGCGCGCGCGCGCGCGCGCG 420
Db 361 GGCACCTGCAACTTCCCGCGAGATCACTCTGCGAGCGCGCGCGCGCGCGCGCGCGCG 420
Qy 421 GGGCGGCGAGATCAAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 480
Db 421 GGGCGGCGAGATCAAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 480
Qy 481 ATGAGCTGCGCGCGCAAGTGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 481 ATGAGCTGCGCGCGCAAGTGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Qy 541 GTGCGCGAGTACGACCGAGATCTGATCGAGATCTGCGGCGAGGCGCGCGCGCGCGCG 600
Db 541 GTGCGCGAGTACGACCGAGATCTGATCGAGATCTGCGGCGAGGCGCGCGCGCGCGCG 600
Qy 601 CTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
```



XX 05-JUL-2001; 2001US-00899575.  
 XX 05-JUL-2000; 2000US-00610313.  
 XX (MEGE/) MEDEDE J Z.  
 XX (BARN/) BARNETT S W.  
 XX (ENGE/) ENGELBRECHT S.  
 XX (RENS/) RENSBERG E J V.  
 XX Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;  
 XX WPI: 2004-060515/06.  
 XX New expression cassette comprising a polynucleotide sequence encoding an  
 XX HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
 XX immunization, generating of packaging cell lines or in producing HIV Type  
 XX C proteins.  
 XX Claim 1; SEQ ID NO 31; 160pp; English.  
 XX The invention relates to an expression cassette comprising a  
 XX polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
 XX also relates to a recombinant expression system for use in a host cell  
 XX comprising an expression cassette, where the polynucleotide sequence  
 XX further comprises control elements capable of driving expression in the  
 XX selected host cell, a cell comprising an expression cassette where the  
 XX polynucleotide sequence further comprises control elements compatible  
 XX with the expression in the cell and a composition for generating an  
 XX immunological response, comprising an expression cassette. The expression  
 XX cassette and the methods of the invention are useful in eliciting an  
 XX immune response, in DNA immunisation, in generation of packaging cell  
 XX lines and in producing HIV Type C proteins. This sequence represents an  
 XX HIV-1 polynucleotide of the invention.  
 XX Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;  
 Query Match 99.2%; Score 2436.2; DB 12; Length 2463;  
 Best Local Similarity 99.6%; Pred. No. 3.6e-293;  
 Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
 1 GTGACGCCACCATGCGCGAGGCCATGAGCAGGCCACACGCGCCCAATCTCTGATGCAG 60  
 1 GTGACGCCACCATGCGCGAGGCCATGAGCAGGCCACACGCGCCCAATCTCTGATGCAG 60  
 61 CGCAGCAACTTCAAGGSCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGC 120  
 61 CGCAGCAACTTCAAGGSCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGC 120  
 121 CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
 121 CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
 181 GGCACACAGATGAAGAGTGTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 240  
 181 GGCACACAGATGAAGAGTGTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 240  
 241 TTCCCGCAGGAGGCGCGGAGTTCCTCCAGCGAGCAGAACCGCGCCCAAGCGCCCAAC 300  
 241 TTCCCGCAGGAGGCGCGGAGTTCCTCCAGCGAGCAGAACCGCGCCCAAGCGCCCAAC 300  
 301 AGCGCGAGTGTGAGGTGCGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGCGCGCGCG 360  
 301 AGCGCGAGTGTGAGGTGCGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGCGCGCGCG 360  
 361 GGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCG 420  
 361 GGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCG 420  
 421 GCGGCGCAGATCAAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 421 GCGGCGCAGATCAAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480

QY 481 ATGAGCCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540  
 DB 481 ATGAGCCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540  
 QY 541 GTGCGCCAGTACGACACAGATCTGATCGAGATCTGCGCGCAGAGAGGCCATCGGACACCGTG 600  
 DB 541 GTGCGCCAGTACGACACAGATCTGATCGAGATCTGCGCGCAGAGAGGCCATCGGACACCGTG 600  
 QY 601 CTGATCGCGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 DB 601 CTGATCGCGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 QY 661 ACCGTGAATCTTCCCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGATG 720  
 DB 661 ACCGTGAATCTTCCCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGATG 720  
 QY 721 GAGGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
 DB 721 GAGGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
 QY 781 ATCTGCGAGGAGTGGAGAGGAGGCAAGATCAACCAAGATCGCGCGCCCGGAAACCCCTAC 840  
 DB 781 ATCTGCGAGGAGTGGAGAGGAGGCAAGATCAACCAAGATCGCGCGCCCGGAAACCCCTAC 840  
 QY 841 AACACCCCGTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 DB 841 AACACCCCGTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
 QY 901 TTCCGCGAGTGAACAG 960  
 DB 901 TTCCGCGAGTGAACAG 960  
 QY 961 CCGCGCGCGCTGAAAG 1020  
 DB 961 CCGCGCGCGCTGAAAG 1020  
 QY 1021 AGCGTGCCTTGAAG 1080  
 DB 1021 AGCGTGCCTTGAAG 1080  
 QY 1081 AACGAGACCCCGCGCATCCGCTACCAAGTGAACAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 DB 1081 AACGAGACCCCGCGCATCCGCTACCAAGTGAACAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 QY 1141 CCAGAGCATCTTCCAGAGCAGATGACCAAGATCTGCGAGCGCTTCCGCGCGCGCGCAAC 1200  
 DB 1141 CCAGAGCATCTTCCAGAGCAGATGACCAAGATCTGCGAGCGCTTCCGCGCGCGCGCAAC 1200  
 QY 1201 GAGATCGTGTATCTACCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260  
 DB 1201 GAGATCGTGTATCTACCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260  
 QY 1261 CGCGCCAGATCGAGAGCTGCGCAGAGCAGCTGCTGCGCTGGGGCTTCCACACCCCGGAC 1320  
 DB 1261 CGCGCCAGATCGAGAGCTGCGCAGAGCAGCTGCTGCGCTGGGGCTTCCACACCCCGGAC 1320  
 QY 1321 AAGAGCAGCAG 1374  
 DB 1321 AAGAGCAGCAG 1374  
 QY 1375 TGACCGTGTGAGCG 1434  
 DB 1375 TGACCGTGTGAGCG 1434  
 QY 1435 AAGCTGTGGGCGAGCTGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1494  
 DB 1435 AAGCTGTGGGCGAGCTGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1494  
 QY 1495 CTGTGCAAGTGTGTGCG 1554  
 DB 1495 CTGTGCAAGTGTGTGCG 1554  
 QY 1501 CTGTGCAAGTGTGTGCG 1560  
 DB 1501 CTGTGCAAGTGTGTGCG 1560  
 QY 1555 GCGAGCTGAGCTGGCGCGAG 1614  
 DB 1555 GCGAGCTGAGCTGGCGCGAG 1614



Db 1561 GCGAGCTGGAGCTGCGCGAGAACCGCGAGATCTCTGCGGAGCCCGTGCACGGCGTGTAC 1620  
Qy 1615 TACGACCCACGACGAGCCTGCTGGCCGAGATCTCAGAAAGCAGGGCCACGACCATGTGGACC 1674  
Db 1621 TACGACCCACGACGAGCCTGCTGGCCGAGATCTCAGAAAGCAGGGCCACGACCATGTGGACC 1680  
Qy 1675 TACGAGATCTACCGAGGCGCTTCAAGAACCTGAGAGCGGCGACAGTACGCCAAGATGGC 1734  
Db 1681 TACGAGATCTACCGAGGCGCTTCAAGAACCTGAGAGCGGCGACAGTACGCCAAGATGGC 1740  
Qy 1735 ACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATGCCATGGAG 1794  
Db 1741 ACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATGCCATGGAG 1800  
Qy 1795 AGCATGCTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTATCCAGAGAGACCTGG 1854  
Db 1801 AGCATGCTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTATCCAGAGAGACCTGG 1860  
Qy 1855 GAGACCTGGTGACCGACTACTGCGCAGCGCCACCTGGATCCCGAGTGGAGTTCTGTGAC 1914  
Db 1861 GAGACCTGGTGACCGACTACTGCGCAGCGCCACCTGGATCCCGAGTGGAGTTCTGTGAC 1920  
Qy 1915 ACCCCCCCTGCTGAAGCTGTGTACCACTGTGAGAAAGAGGCCCATCATCGGGCCGAG 1974  
Db 1921 ACCCCCCCTGCTGAAGCTGTGTACCACTGTGAGAAAGAGGCCCATCATCGGGCCGAG 1980  
Qy 1975 ACCTTCTACGTGAGCGGCGGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2034  
Db 1981 ACCTTCTACGTGAGCGGCGGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2040  
Qy 2035 ACCGACCGGGCCGCGACAGATCTGTAGCTGTGACCGAGACCAACCAAGAGACCGAG 2094  
Db 2041 ACCGACCGGGCCGCGACAGATCTGTAGCTGTGACCGAGACCAACCAAGAGACCGAG 2100  
Qy 2095 CTGACGGGCATCCAGCTGGCCCTGCGAGACAGCGGCGAGAGTGAACATCGTGACCGAC 2154  
Db 2101 CTGACGGGCATCCAGCTGGCCCTGCGAGACAGCGGCGAGAGTGAACATCGTGACCGAC 2160  
Qy 2155 AGCCAGTACGGCTGGGATCATCATCAGCGCCAGCCCGACAGAGCGAGGAGTGTGTG 2214  
Db 2161 AGCCAGTACGGCTGGGATCATCATCAGCGCCAGCCCGACAGAGCGAGGAGTGTGTG 2220  
Qy 2215 AACCAAGATCATGACGACGCTGATCAAGAGGAGAGTGTACCTGAGCTGGGTGCCGCC 2274  
Db 2221 AACCAAGATCATGACGACGCTGATCAAGAGGAGAGTGTACCTGAGCTGGGTGCCGCC 2280  
Qy 2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGCAAG 2334  
Db 2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGCAAG 2340  
Qy 2335 GTGCTGTTCTGAGCGGATCGATCGGCGGATCTGTGATCTTACAGTACATGAGACGACCTG 2394  
Db 2341 GTGCTGTTCTTGGAGCGCATCGATGGCGGATCTGTGATCTTACAGTACATGAGACGACCTG 2400  
Qy 2395 TACGTGGGACGGCGGCGCTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGTGAA 2454  
Db 2401 TACGTGGGACGGCGGCGCTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGTGAA 2460  
Qy 2455 TTC 2457  
Db 2461 TTC 2463

RESULT 8  
ADCl3231  
ID ADCl3231 standard; DNA; 3930 BP.

XX ADCl3231;

XX 18-DEC-2003 (first entry)

DE DNA of HIV construct GagComplPolmutAtt\_C SEQ ID NO 10.

XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX Human immunodeficiency virus.  
OS WO2003004620-A2.  
XX 16-JAN-2003.  
XX 05-JUL-2002; 2002WO-US021420.  
XX 05-JUL-2001; 2001US-0303192P.  
XX 31-AUG-2001; 2001US-0316860P.  
XX 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ;  
XX WPI; 2003-221593/21.  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX Disclosure; Fig 7; 301pp; English.  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intradermally, intravenously, intranasally,  
CC subcutaneously, intrarectally, or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
SQ Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;  
Query Match 99.1%; Score 2434.8; DB 10; Length 3930;  
Best Local Similarity 99.9%; Pred. No. 5.1e-293;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 14 TGGCCGAGGCCATGAGCCAGGCCACCGCGCAACATCTGATGCGCGCAGCAACTTCA 73  
Db 1487 TCGCGAGGCCATGAGCCAGGCCACCGCGCAACATCTGATGCGCGCAGCAACTTCA 1546  
Qy 74 AGGCCCCCAAGCGCATCATCAAGTGTCTCAATCGCGCAAGGAGGCCACATCGCCCCGA 133  
Db 1547 AGGCCCCCAAGCGCATCATCAAGTGTCTCAATCGCGCAAGGAGGCCACATCGCCCCGA 1606  
Qy 134 ACTCGCGGCCCGCCGCGAGAGGCTGCTGGAGTGGCGCAAGGAGGCCACCGAGTGA 193  
Db 1607 ACTCGCGGCCCGCCGCGAGAGGCTGCTGGAGTGGCGCAAGGAGGCCACCGAGTGA 1666  
Qy 194 AGGACTGCGAGCGCGCGAGGCCAACTTCTTCCGCGAGGAGCTGGCTTCCCGCCAGGGCA 253  
Db 1667 AGGACTGCGAGCGCGCGAGGCCAACTTCTTCCGCGAGGAGCTGGCTTCCCGCCAGGGCA 1726  
Qy 254 AGGCCCCGAGTTCCTCCAGCGAGAGAACCCCGCGCAACAGCCCCCAAGCGCGAGCTGC 313  
Db 1727 AGGCCCCGAGTTCCTCCAGCGAGAGAACCCCGCGCAACAGCCCCCAAGCGCGAGCTGC 1786  
Qy 314 AGGTGCGGGCGCAACACCCCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
Db 1787 AGGTGCGGGCGCAACACCCCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
Qy 374 TCCCCAGATCACCTGTGGCAGCGGCCCTTGGTGAGCATCAAGGTGGCGCGCGCGAGATCA 433

```
Db 1847 TCCCCAGATCACCTGTGGCAGGCCCCCTGGTGAGCATCAAGGTGGCGCGCAGATCA 1906
Qy 434 AGGAGGCCCTGCTGAGACACCGGCGCGACACCGTGTCTGGAGGAGATGAGCTTGCCCG 493
Db 1907 AGGAGGCCCTGCTGAGCTTCGGCTCGGCGCGACACCGTGTCTGGAGGAGATGAGCTTGCCCG 1966
Qy 494 GCAAGTGGAAAGCCCAAGATGATCGGCGCATCGGCGGCTTCAATCAAGGTGGCGCAGATCG 553
Db 1967 GCAAGTGGAAAGCCCAAGATGATCGGCGCATCGGCGGCTTCAATCAAGGTGGCGCAGATCG 2026
Qy 554 ACCAGATCCTCATCGAGATCTGCGCAAGAGGCGCATCGGCAACGCTGATCGGCCCA 613
Db 2027 ACCAGATCCTCATCGAGATCTGCGCAAGAGGCGCATCGGCAACGCTGATCGGCCCA 2086
Qy 614 CCCCCTGGAACATCATCGGCGCAACATGCTGAACCCAGCTGGGTGCAACCTGGAACCTTCC 673
Db 2087 CCCCCTGGAACATCATCGGCGCAACATGCTGAACCCAGCTGGGTGCAACCTGGAACCTTCC 2146
Qy 674 CCATCAGCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGCATGGAAGCGCCCAAGG 733
Db 2147 CCATCAGCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGCATGGAAGCGCCCAAGG 2206
Qy 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793
Db 2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266
Qy 794 TGAAGAGGAGGCGCAAGATCAACCAAGATCGGCGCCCGAGAACCCCTACAACACCCCGTGT 853
Db 2267 TGAAGAGGAGGCGCAAGATCAACCAAGATCGGCGCCCGAGAACCCCTACAACACCCCGTGT 2326
Qy 854 TCGCCATCAAGAAGAGGAGACGACCAAGTGGCGCAAGCTGGTGACTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAAGAGGAGACGACCAAGTGGCGCAAGCTGGTGACTTCCGCGAGCTGA 2386
Qy 914 ACAGCGCACCCAGACTTCTGGAGGTGAGCTGGGCAATCCCCCAACCGCGCGGCTGA 973
Db 2387 ACAGCGCACCCAGACTTCTGGAGGTGAGCTGGGCAATCCCCCAACCGCGCGGCTGA 2446
Qy 974 AGAAGAGAGAGCGTGACCGTGTGGACGTGGCGAGCGCTACTTACGCGTGCCCTGG 1033
Db 2447 AGAAGAGAGAGCGTGACCGTGTGGACGTGGCGAGCGCTACTTACGCGTGCCCTGG 2506
Qy 1034 ACAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGATCAACAAAGAGACCCCG 1093
Db 2507 ACAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGATCAACAAAGAGACCCCG 2566
Qy 1094 GCATCCGCTACAGTACAAAGTGTGCCCCAGGCTGGAGGGCAGCGCCAGCATCTTCC 1153
Db 2567 GCATCCGCTACAGTACAAAGTGTGCCCCAGGCTGGAGGGCAGCGCCAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCGCTTCCGCGCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCCTGGAGCGCTTCCGCGCCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACCAGGCCCCCTGTACGTGGGCGAGCAGCTGGAGATCGGCGAGCAGCGCCCAAGATCG 1273
Db 2687 ACCAGGCCCCCTGTACGTGGGCGAGCAGCTGGAGATCGGCGAGCAGCGCCCAAGATCG 2746
Qy 1274 AGGAGCTGGCGAAGCACTGTGCGCTGGGGCTTCAACACCCCGCAAGAGCAACAGA 1333
Db 2747 AGGAGCTGGCGAAGCACTGTGCGCTGGGGCTTCAACACCCCGCAAGAGCAACAGA 2806
Qy 1334 AGGAGCCCCCTTCTGCCCATCGAGCTGCAACCCCGCAAGTGGACCGTGCAGGCCATCG 1393
Db 2807 AGGAGCCCCCTTCTGCCCATCGAGCTGCAACCCCGCAAGTGGACCGTGCAGGCCATCG 2866
Qy 1394 AGCTGCCCGAAGAGAGCTGGACCGCTGAACGACATCCAGAACTGGTGGGCAAGCTGA 1453
Db 2867 AGCTGCCCGAAGAGAGCTGGACCGCTGAACGACATCCAGAACTGGTGGGCAAGCTGA 2926
Qy 1454 ACTGGGCGAGCAGATCTTACCCCGGCATCAAGGTGGCGAGCTGTGAAGCTGTGCGCG 1513
```

```
Db 2927 ACTGGGCGAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCG 2986
Qy 1514 GCGCCAAGGCCCTGACCGACATCGTGGCCCTTGACCGGAGGCGGAGCTGAGCTGGCGG 1573
Db 2987 GCGCCAAGGCCCTGACCGACATCGTGGCCCTTGACCGGAGGCGGAGCTGAGCTGGCGG 3046
Qy 1574 AGAACCGGAGATCCTCGCGAGGCCCTGCACGCGGTGTACTACGACCCGAGCAAGACC 1633
Db 3047 AGAACCGGAGATCCTCGCGAGGCCCTGCACGCGGTGTACTACGACCCGAGCAAGACC 3106
Qy 1634 TGGTGGCCGAGATCCAGAAGCAGGCGCACCAAGTGGACCTTACAGATCTTACAGGAGC 1693
Db 3107 TGGTGGCCGAGATCCAGAAGCAGGCGCACCAAGTGGACCTTACAGATCTTACAGGAGC 3166
Qy 1694 CTTTCAAGAACTCTGAAGACCGGCAAGTACGCAAGATGCGCACGCCCCACACCAACGACG 1753
Db 3167 CTTTCAAGAACTCTGAAGACCGGCAAGTACGCAAGATGCGCACGCCCCACACCAACGACG 3226
Qy 1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGGAGAGCATCGTGAATCTGGGGCA 1813
Db 3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGGAGAGCATCGTGAATCTGGGGCA 3286
Qy 1814 AGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGCGAGACCTGTGGACCGACT 1873
Db 3287 AGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGTGGAGACCTGTGGACCGACT 3346
Qy 1874 ACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCTGTGAAGC 1933
Db 3347 ACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCTGTGAAGC 3406
Qy 1934 TGTGGTACCAAGCTGGAGAGGAGGCCATCATCGGCGCGGAGACCTTCTACGTGAAGCGG 1993
Db 3407 TGTGGTACCAAGCTGGAGAGGAGGCCATCATCGGCGCGGAGACCTTCTACGTGAAGCGG 3466
Qy 1994 CCGCAACCCGAGACCAAGATCGGCAAGCGCGCTAGCTGACCGAGCGGCGCGGCGAGA 2053
Db 3467 CCGCAACCCGAGACCAAGATCGGCAAGCGCGCTAGCTGACCGAGCGGCGCGGCGAGA 3526
Qy 2054 AGATCGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTGCAGGCCATCCAGCTGG 2113
Db 3527 AGATCGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTGCAGGCCATCCAGCTGG 3586
Qy 2114 CCTTGCAGGACAGCGGCGAGCGAGGTGAACATCGTGACCGAGCAGCAGTACGCCCTGGGCA 2173
Db 3587 CCTTGCAGGACAGCGGCGAGCGAGGTGAACATCGTGACCGAGCAGCAGTACGCCCTGGGCA 3646
Qy 2174 TCATCCAGGCCACCGCGACAGAGCGAGCGAGCTGGAACAGATCATCGAGCAGC 2233
Db 3647 TCATCCAGGCCACCGCGACAGAGCGAGCGAGCTGGAACAGATCATCGAGCAGC 3706
Qy 2234 TGATCAAGAAAGGAGAGAGGTGTACTGTAGCTGGGTGCCCGCCCAAGGGCATCGGCGGCA 2293
Db 3707 TGATCAAGAAAGGAGAGAGGTGTACTGTAGCTGGGTGCCCGCCCAAGGGCATCGGCGGCA 3766
Qy 2294 ACAGCAGATCGCAAGCTGGTGAAGGCGCATCCGCAAGGTGCTGTCTTGAAGCGCA 2353
Db 3767 ACAGCAGATCGCAAGCTGGTGAAGGCGCATCCGCAAGGTGCTGTCTTGAAGCGCA 3826
Qy 2354 TCGATGGCGGCATCGTGAATCTACAGTACATGAGCAGCTGTAGCTGGGCGAGCGCGGCC 2413
Db 3827 TCGATGGCGGCATCGTGAATCTACAGTACATGAGCAGCTGTAGCTGGGCGAGCGCGGCC 3886
Qy 2414 CTAGGATCGAATTAAGCTTCCCGGGCTAGCACCGGT 2451
Db 3887 CTAGGATCGAATTAAGCTTCCCGGGCTAGCACCGGT 3924
```

RESULT 9

ADCL1322

ID ADCL13232 standard; DNA; 3930 BP.

XX

AC

ADCL13232;

XX



QY	1454	ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGTCAAGCTGCTGGCG	1513
DB	2927	ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGTCAAGCTGCTGGCG	2986
QY	1514	GGCCAAAGGCCCTGACCGACATCTGTGCCCCTTGACCCGAGGAGGCCGAGCTGGAGCTGGCCG	1573
DB	2987	GGCCAAAGGCCCTGACCGACATCTGTGCCCCTTGACCCGAGGAGGCCGAGCTGGAGCTGGCCG	3046
QY	1574	AGAACCCGGAGATCCTGGCGAGCCCGTGACCGCGGTGTACTACGACCCAGACGAGACG	1633
DB	3047	AGAACCCGGAGATCCTGGCGAGCCCGTGACCGCGGTGTACTACGACCCAGACGAGACG	3106
QY	1634	TGTTGGCCGAGATCCAGAAGCAGGGCCACGACCAAGTGGACTTACCAGATCTTACCAGGAGC	1693
DB	3107	TGTTGGCCGAGATCCAGAAGCAGGGCCACGACCAAGTGGACTTACCAGATCTTACCAGGAGC	3166
QY	1694	CTTTCAAGAACTTGAAGACCGCGAAGTACGCGCAAGATGCGGACCGGCCACACCAACGAGC	1753
DB	3167	CTTTCAAGAACTTGAAGACCGCGAAGTACGCGCAAGATGCGGACCGGCCACACCAACGAGC	3226
QY	1754	TGAAGCAGCTGACCGAGGCCGTGAGAAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA	1813
DB	3227	TGAAGCAGCTGACCGAGGCCGTGCGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCA	3286
QY	1814	AGACCCCAAAGTTCCGCTGCCATCAGAAAGAGACTCTGGAGACTGTGTGACCGACT	1873
DB	3287	AGACCCCAAAGTTCCGCTGCCATCAGAAAGAGACTCTGGAGACTGTGTGACCGACT	3346
QY	1874	ACTGGCAGGCCACTGTGATCCCGAGTGGAGTTCTGTGAACACCCCGCCCTGGTGAAGC	1933
DB	3347	ACTGGCAGGCCACTGTGATCCCGAGTGGAGTTCTGTGAACACCCCGCCCTGGTGAAGC	3406
QY	1934	TGTGGTACCAGCTGGAGAAGAGCCCATCATCGCGCCCGAGACTTCTACGTGTGAACGCGC	1993
DB	3407	TGTGGTACCAGCTGGAGAAGAGCCCATCATCGCGCCCGAGACTTCTACGTGTGAACGCGC	3466
QY	1994	CCGCCAACCGCGAGACCAAGATCGCGAAGCGCGGCTACGTGACCGACCGGGCCGGGAGA	2053
DB	3467	CCGCCAACCGCGAGACCAAGATCGCGAAGCGCGGCTACGTGACCGACCGGGCCGGGAGA	3526
QY	2054	AGATCGTGAGCTGACCGAGACCAACACAGAAAGACCGAGCTGCGAGCCATCCAGCTGG	2113
DB	3527	AGATCGTGAGCTGACCGAGACCAACACAGAAAGACCGAGCTGCGAGCCATCCAGCTGG	3586
QY	2114	CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA	2173
DB	3587	CCCTGCAGGACAGCGGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCA	3646
QY	2174	TCAATCCAGGCCAGCCCGACAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGC	2233
DB	3647	TCAATCCAGGCCAGCCCGACAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGC	3706
QY	2234	TGATCAAGAAAGGAGAGGTGTACTGTGAGCTGGGTGCCCGCCCAAGGGGCATCGGGGCA	2293
DB	3707	TGATCAAGAAAGGAGAGGTGTACTGTGAGCTGGGTGCCCGCCCAAGGGGCATCGGGGCA	3766
QY	2294	ACGAGCAGATCGCAACAGCTGGTGAACAGGGCATCCGCAAGGTGCTGTTCTCGACGGCA	2353
DB	3767	ACGAGCAGATCGCAACAGCTGGTGAACAGGGCATCCGCAAGGTGCTGTTCTCGACGGCA	3826
QY	2354	TCCATGGCGGCATCGTGATCTTACAGTATGACGACCTGTATCGTGGGAGCGGGGCC	2413
DB	3827	TCCATGGCGGCATCGTGATCTTACAGTATGACGACCTGTATCGTGGGAGCGGGGCC	3886
QY	2414	CTAGGATCGAATTAAGCTTCCCGGGGCTTAGCACCGGT	2451
DB	3887	CTAGGATCGAATTAAGCTTCCCGGGGCTTAGCACCGGT	3924

RESULT 10  
ACA03591  
ID ACA0

XX	ACA03591;
AC	
XX	
XX	22-MAY-2003 (first entry)
DT	
XX	
XX	Synthetic DNA encoding immunogenic HIV peptide #74.
DE	
XX	
XX	Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW	gene therapy; packaging cell line; humoral immune response;
KW	cellular immune response; gene delivery vector; DNA immunisation; ds.
XX	
XX	Synthetic.
OS	
XX	
FN	WO2003004657-A1.
XX	
XX	16-JAN-2003.
PD	
XX	
XX	05-JUL-2002; 2002WO-US021421.
PF	
XX	
XX	05-JUL-2001; 2001US-0303192P.
PR	
PR	31-AUG-2001; 2001US-0316860P.
PR	16-JAN-2002; 2002US-0349728P.
PR	16-JAN-2002; 2002US-0349733P.
PR	16-JAN-2002; 2002US-0349871P.
XX	
PA	(CHIR ) CHIRON CORP.
XX	
XX	
PI	Zur Megede J, Barnett SW, Lian Y;
XX	
XX	WPI; 2003-221602/21.
DR	
XX	
PT	New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT	polypeptides, useful as immunogenic compositions or vaccines for
PT	generating humoral or cellular immune responses against HIV in a subject,
PT	especially humans.
PS	
XX	Example 1; Fig 79; 262pp; English.
PS	
XX	
CC	The invention describes a synthetic polynucleotide encoding 2 or more
CC	immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC	derived from different HIV subtypes. The polynucleotide is useful for
CC	immunisation, generation of packaging cell lines, or production of HIV
CC	polypeptides. The polynucleotide and its encoded proteins are useful as
CC	immunogenic compositions or vaccines for generating humoral or cellular
CC	immune responses against HIV in a subject, or for inducing neutralising
CC	antibodies against HIV. The gene delivery vector comprising the
CC	polynucleotide is also useful for DNA immunisation of, or for generating
CC	an immune response (e.g. a humoral or cellular immune response) in, a
CC	subject such as a mammal, particularly a human. This sequence encodes a
CC	human immunodeficiency virus immunogenic peptide
XX	
XX	Sequence 5184 BP: 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

	Query Match	99.1%;	Score 2434.8;	DB 8;	Length 5184;
	Best Local Similarity	99.9%;	Pred. No. 4.9e-293;		
	Matches 2436;	Conservative	0;	Mismatches 2;	Indels 0;
					Gaps 0;
Qy	14	TGGCCGAGGCGCATGAGCCAGGCCACAGCGCCCAACATCTCTGATGCGAGCGAGCAACTTCA	73		
Db	2741	TGCGCGGNGGCATGAGCCAGGCCACAGCGCCCAACATCTGTGTCAGCGCGAGCACTTCA	2800		
Qy	74	AGGGCCCCAAGCGGATCATCAAGTGTCTTAACTTGCGGCAAGAGGGGCCACATCGCCCCGCA	133		
Db	2801	AGGGCCCCAAGCGCATCATCAAGTGTCTTAACTTGCGGCAAGAGGGGCCACATCGCCCCGCA	2860		
Qy	134	ACTTCCGCGCCCCCGCCAGAGGGGTGCTTGGAAAGTGCGGCAAGGAGGGGCCACCAGATGA	193		
Db	2861	ACTTCCGCGCCCCCGCCAGAGGGGTGCTTGGAAAGTGCGGCAAGGAGGGGCCACCAGATGA	2920		
Qy	194	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGGCGCTTCCCCCAGGGCA	253		
Db	2921	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGGCGCTTCCCCCAGGGCA	2980		

QY 254 AGGCGCGGAGTTCCAGCGAGCAGAAACCGCGCAACAGCGCCACACGCGCGAGCTGC 313  
 Db |||||  
 QY 2981 AGGCGCGGAGTTCCCGAGGAGCAGAGACCGCGCAACAGCGCCACACGCGCGAGCTGC 3040  
 Db |||||  
 QY 314 AGGTGCGGCGCAGCAACCCCGCAGCGAGGCGCGCGCGAGCGCCAGGGCACCTTGAACT 373  
 Db |||||  
 QY 3041 AGGTGCGGCGCAGCAACCCCGCAGCGAGGCGCGCGCGAGCGCCAGGGCACCTTGAACT 3100  
 Db |||||  
 QY 374 TCCCGCCAGATCACTGTGGCAGCGCCCGCTGGTGAGGATCAAGGTGGCGGCGCAGATCA 433  
 Db |||||  
 QY 3101 TCCCGCCAGATCACTGTGGCAGCGCCCGCTGGTGAGGATCAAGGTGGCGGCGCAGATCA 3160  
 Db |||||  
 QY 434 AGGAGGCGCTGTGGACACCGCGCGCGCAGCAGACCGCTGTGGAGGAGATGAGCTGCGCG 493  
 Db |||||  
 QY 3161 AGGAGGCGCTGTGGCCACCGCGCGCGCAGCAGACCGCTGTGGAGGAGATGAGCTGCGCG 3220  
 Db |||||  
 QY 494 GCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTCATCAAGGTGGCGGCCAGTAGC 553  
 Db |||||  
 QY 3221 GCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTCATCAAGGTGGCGGCCAGTAGC 3280  
 Db |||||  
 QY 554 ACCAGATCTGTGATCGAGATCTGGCGCAGAGAGGCCATCGGCGACCGTGCTGATCGGCGCGCA 613  
 Db |||||  
 QY 3281 ACCAGATCTGTGATCGAGATCTGGCGCAGAGAGGCCATCGGCGACCGTGCTGATCGGCGCGCA 3340  
 Db |||||  
 QY 614 CCGCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCACTTGAACCTTCC 673  
 Db |||||  
 QY 3341 CCGCGGTGNAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCACTTGAACCTTCC 3400  
 Db |||||  
 QY 674 CCATCAGCGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGG 733  
 Db |||||  
 QY 3401 CCATCAGCGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGG 3460  
 Db |||||  
 QY 734 TGAAGCAGTGGCGCTGACCGAGGAGAGATCAAGGCGCTGACCGCGCATCTCGGAGGAGA 793  
 Db |||||  
 QY 3461 TGAAGCAGTGGCGCTGACCGAGGAGAGATCAAGGCGCTGACCGCGCATCTCGGAGGAGA 3520  
 Db |||||  
 QY 794 TGAAGAGGAGGCGCAAGATCAACAGATCGGCGCGCGAGAACCTTACAAACACCCCGGT 853  
 Db |||||  
 QY 3521 TGAAGAGGAGGCGCAAGATCAACAGATCGGCGCGCGAGAACCTTACAAACACCCCGGT 3580  
 Db |||||  
 QY 854 TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGACTTTCGCGAGCTGA 913  
 Db |||||  
 QY 3581 TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGACTTTCGCGAGCTGA 3640  
 Db |||||  
 QY 914 ACAAGCGCACCCAGGACTTCTGGAGGTGACCTGGGATCCCCACCGCGCGCGCTGA 973  
 Db |||||  
 QY 3641 ACAAGCGCACCCAGGACTTCTGGAGGTGACCTGGGATCCCCACCGCGCGCGCTGA 3700  
 Db |||||  
 QY 974 AGAAGAGAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTTTCAGCGTGGCCCTGG 1033  
 Db |||||  
 QY 3701 AGAAGAGAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTTTCAGCGTGGCCCTGG 3760  
 Db |||||  
 QY 1034 ACAGGACTTTCGCAAGTACACCGCTTTCACATCCCCAGCATCAACACAGAGACCCCG 1093  
 Db |||||  
 QY 3761 ACAGGACTTTCGCAAGTACACCGCTTTCACATCCCCAGCATCAACACAGAGACCCCG 3820  
 Db |||||  
 QY 1094 GCATCGCTACAGTACAGTGTGCTGCGCGCGGCTGGAAGGCGAGCGCCAGCATCTTCC 1153  
 Db |||||  
 QY 3821 GCATCGCTACAGTACAGTGTGCTGCGCGCGGCTGGAAGGCGAGCGCCAGCATCTTCC 3880  
 Db |||||  
 QY 1154 AGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCTGATCT 1213  
 Db |||||  
 QY 3881 AGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCTGATCT 3940  
 Db |||||  
 QY 1214 ACCAGGCGCCCTGTGATGGGAGCGACTTGGAGATCGGCGAGCAGCCGCGCAAGATCG 1273  
 Db |||||  
 QY 3941 ACCAGGCGCCCTGTGATGGGAGCGACTTGGAGATCGGCGAGCAGCCGCGCAAGATCG 4000  
 Db |||||  
 QY 1274 AGGAGCTGGCGAAGCACCTGCTGCGCTGGGGCTTACCAACCCCGCAGAGAGCACCA 1333  
 Db |||||  
 QY 4001 AGGAGCTGGCGAAGCACCTGCTGCGCTGGGGCTTACCAACCCCGCAGAGAGCACCA 4060  
 Db |||||  
 QY 1334 AGGAGCGCCCTTCTGCGCCATCGAGCTGCACCGCGCAAGTGGAGCGGTGACGCGCCATCG 1393  
 Db |||||

Db 4061 AGGAGCGCCCTTCTGCGCCATCGAGCTGCACCCGCAACAGTGGACCGTGAGCCATCG 4120  
 QY |||||  
 QY 1394 AGCTGCCGAGAGAGAGAGCTGGACCGTGAAACGACATCCAGAGCTGGTGGCAAGCTGA 1453  
 Db |||||  
 QY 4121 AGCTGCCGAGAGAGAGAGCTGGACCGTGAAACGACATCCAGAGCTGGTGGCAAGCTGA 4180  
 Db |||||  
 QY 1454 ACTGGGCGAGCAGATCTACCCCGGCGATCAAGGTGGCGCAGCTGTGCAAGCTGTCTGCGCG 1513  
 Db |||||  
 QY 4181 ACTGGGCGAGCAGATCTACCCCGGCGATCAAGGTGGCGCAGCTGTGCAAGCTGTCTGCGCG 4240  
 Db |||||  
 QY 1514 GCGCAAGGCGCTGACCGGATCTGTCCTTGAACGAGGAGCGGAGCTGGAGCTGGCGCG 1573  
 Db |||||  
 QY 4241 GCGCAAGGCGCTGACCGGATCTGTCCTTGAACGAGGAGCGGAGCTGGAGCTGGCGCG 4300  
 Db |||||  
 QY 1574 AGAACCGGAGATCTTGGCGAGCGCGCTGACGCGGCTGTACTACACCCCGAGCAAGGACC 1633  
 Db |||||  
 QY 4301 AGAACCGGAGATCTTGGCGAGCGCGCTGACGCGGCTGTACTACACCCCGAGCAAGGACC 4360  
 Db |||||  
 QY 1634 TGGTGGCGGAGATCCAGAGCAGGCGCCACGACGAGTGGACCTTACCAGATCTTACCAGGAGC 1693  
 Db |||||  
 QY 4361 TGGTGGCGGAGATCCAGAGCAGGCGCCACGACGAGTGGACCTTACCAGATCTTACCAGGAGC 4420  
 Db |||||  
 QY 1694 CTTTCAAGAACCTTGAAGACCGGCAAGTACGCGCAAGTACGCGCAAGTGGCGCACCGCCACCAACGAGC 1753  
 Db |||||  
 QY 4421 CTTTCAAGAACCTTGAAGACCGGCAAGTACGCGCAAGTGGCGCACCGCCACCAACGAGC 4480  
 Db |||||  
 QY 1754 TGAAGCAGCTGACCGAGGCGCTGCAAGAGATCGCCATGGAGAGCATCTGTGATCTGGGGCA 1813  
 Db |||||  
 QY 4481 TGAAGCAGCTGACCGAGGCGCTGCAAGAGATCGCCATGGAGAGCATCTGTGATCTGGGGCA 4540  
 Db |||||  
 QY 1814 AGACCCCAAGTTCGCGCTGCCATCCAGAGAGGAGACCTGGGAGACCTGTGGACCGACT 1873  
 Db |||||  
 QY 4541 AGACCCCAAGTTCGCGCTGCCATCCAGAGAGGAGACCTGGGAGACCTGTGGACCGACT 4600  
 Db |||||  
 QY 1874 ACTGSGAGGCGCACCTTGGATCCCGAGTGGGAGTTCTGTAACACACCCCGCTGTGTAAGC 1933  
 Db |||||  
 QY 4601 ACTGSGAGGCGCACCTTGGATCCCGAGTGGGAGTTCTGTAACACACCCCGCTGTGTAAGC 4660  
 Db |||||  
 QY 1934 TGTGTTACAGCTGGAGAGGAGCGCATCATCGGCGCGCGAGACCTTTCAGCTGGAGCGCG 1993  
 Db |||||  
 QY 4661 TGTGTTACAGCTGGAGAGGAGCGCATCATCGGCGCGCGAGACCTTTCAGCTGGAGCGCG 4720  
 Db |||||  
 QY 1994 CCGCAACCGCGAGACCAAGATCGCAAGCGCGCTAGCTGACACCGCGGCGCGGAGC 2053  
 Db |||||  
 QY 4721 CCGCAACCGCGAGACCAAGATCGCAAGCGCGCTAGCTGACACCGCGGCGCGGAGC 4780  
 Db |||||  
 QY 2054 AGATCGTGAGCTGACCGGAGACCAACCGAGAGACCGAGCTGCAGGCGCATTCAGCTGG 2113  
 Db |||||  
 QY 4781 AGATCGTGAGCTGACCGGAGACCAACCGAGAGACCGAGCTGCAGGCGCATTCAGCTGG 4840  
 Db |||||  
 QY 2114 CCCTGAGGAGACGCGGAGCGAGGTGAACATCGTGACCGAGCAGCAGCTTCCGCTGGGCA 2173  
 Db |||||  
 QY 4841 CCCTGAGGAGACGCGGAGCGAGGTGAACATCGTGACCGAGCAGCAGCTTCCGCTGGGCA 4900  
 Db |||||  
 QY 2174 TCATCCAGGCGCGCGCGAGAGCGAGCTGTGAAACAGATCAACAGAGCAGC 2233  
 Db |||||  
 QY 4901 TCATCCAGGCGCGCGCGAGAGCGAGCTGTGAAACAGATCAACAGAGCAGC 4960  
 Db |||||  
 QY 2234 TGATCAAGAGGAGAGGTGTACTGAGCTGGTGGCGCGCGCGCAAGGGCGCATTCGCGGCA 2293  
 Db |||||  
 QY 4961 TGATCAAGAGGAGAGGTGTACTGAGCTGGTGGCGCGCGCGCAAGGGCGCATTCGCGGCA 5020  
 Db |||||  
 QY 2294 ACGAGCAGATCGAACAGCTGGTGAAGAGGCGATCCGCAAGGCTGTGTTCTTGGAGCGGCA 2353  
 Db |||||  
 QY 5021 ACGAGCAGATCGAACAGCTGGTGAAGAGGCGATCCGCAAGGCTGTGTTCTTGGAGCGGCA 5080  
 Db |||||  
 QY 2354 TCGATGGCGGAGCTCGTGATCTACAGTACATGAGCAGCTGTGACGTGGCGAGCGCGCGC 2413  
 Db |||||  
 QY 5081 TCGATGGCGGAGCTCGTGATCTACAGTACATGAGCAGCTGTGACGTGGCGAGCGCGCGC 5140  
 Db |||||  
 QY 2414 CTAGATCGATTAAGAGCTTCCGCGGCTAGCACCGGT 2451  
 Db |||||

Db 5141 CTAGATCGATTAAAGCTTCCGGGGCTAGACGGT 5178

RESULT 11

ADCI3279

ID ADCI3279 standard; DNA; 5184 BP.

XX

XX

AC ADCI3279;

XX

XX

DT 18-DEC-2003 (first entry)

XX

XX

DE DNA of HIV construct TatRevWefagCpolIna\_C SEQ ID NO 58.

XX

KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;

KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX

XX

OS Human immunodeficiency virus.

XX

XX

PN WO2003004620-A2.

XX

XX

PD 16-JAN-2003.

XX

XX

PF 05-JUL-2002; 2002WO-US021420.

XX

XX

PR 05-JUL-2001; 2001US-0303192P.

PR

PR 31-AUG-2001; 2001US-0316860P.

PR

PR 16-JAN-2002; 2002US-0349871P.

XX

XX

XX

PA (CHIR ) CHIRON CORP.

PA

PA (UYST-) UNIV STELLENBOSCH.

XX

PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX

XX

XX

DR WPI; 2003-221593/21.

XX

XX

PT New expression cassette comprising a polynucleotide sequence encoding a

PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

PT Prot, or Rev polypeptide, useful for immunization, or generating

PT packaging cell lines.

XX

XX

PS Disclosure; Fig 55; 301pp; English.

XX

XX

CC The invention relates to a novel expression cassette comprising a

CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,

CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel

CC expression cassette can be used to treat HIV type C by gene therapy or

CC used in the development of a vaccine. The gene delivery vector is

CC administered intramuscularly, intracutaneously, intravenously,

CC subcutaneously, intradermally, transdermally, intravaginally,

CC intrarectally, orally or intravenously. The expression cassette is useful

CC for immunisation, generating packaging cell lines and producing HIV

CC polypeptides. This polynucleotide sequence represents the DNA of an HIV

CC Type C related sequence of the invention.

XX

XX

SQ Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

Query Match 99.1%; Score 2434.8; DB 10; Length 5184;

Best Local Similarity 99.9%; Pred. No. 4.9e-293;

Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCGCATGAGCGAGGCCACACGAGCCCAATCTCTGATCGAGCGCAACTTCA 2800

Db

Qy 74 AGGGCCCGAGCGCATCATCAAGTGGCTTCACTGGCGAGGAGGCCACATCGCCGCA 133.

Db

Db 2801 AGGGCCCGAGCGCATCATCAAGTGGCTTCACTGGCGAGGAGGCCACATCGCCGCA 2860

Qy 134 ACTCGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGGCGAGGAGGCCACAGATGA 193

Db

Db 2861 ACTCGCGCGCCCCCGGCAAGAGGGCTGCTGGAAGTGGCGAGGAGGCCACAGATGA 2920

Qy 194 AGGACTGCAKCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGAGGGCA 253

Db 2921 AGGACTGCAKCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGAGGGCA 2980

Qy

254 AGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCAACAGCCCGCGAGTGC 313

Db

2981 AGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCAACAGCCCGCGAGTGC 3040

Qy

314 AGGTGCGCGCGCAACACCCCGCAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373

Db

3041 AGGTGCGCGCGCAACACCCCGCAGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100

Qy

374 TCCCGCAGATCACCTGTGGCAGGCGCCCGCTGTGTGAGCATCAAGTGGCGCGCGCGCG 433

Db

3101 TCCCGCAGATCACCTGTGGCAGGCGCCCGCTGTGTGAGCATCAAGTGGCGCGCGCG 3160

Qy

434 AGGAGGCGCTGTGGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493

Db

3161 AGGAGGCGCTGTGGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3220

Qy

494 GCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGTGGCGCGCG 553

Db

3221 GCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGTGGCGCGCG 3280

Qy

554 ACCAGATCCTGTGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGTGATCGCGCGCG 613

Db

3281 ACCAGATCCTGTGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGTGATCGCGCGCG 3340

Qy

614 CCCCCTGGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGGTGCACCTGTAACCTTCC 673

Db

3341 CCCCCTGGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGGTGCACCTGTAACCTTCC 3400

Qy

674 CCATCAGCCCCCATCGAGACCGTGCCTGAGCTGAAAGCCCGGATGAGAGCGCCCCAAGG 733

Db

3401 CCATCAGCCCCCATCGAGACCGTGCCTGAGCTGAAAGCCCGGATGAGAGCGCCCCAAGG 3460

Qy

734 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCAGAGAGA 793

Db

3461 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCAGAGAGA 3520

Qy

794 TGGAGAGGAGGGCAAGATCACCAAGATCGCGCGCGAGAACCCCTACAACACCCCGTGT 853

Db

3521 TGGAGAGGAGGGCAAGATCACCAAGATCGCGCGCGAGAACCCCTACAACACCCCGTGT 3580

Qy

854 TCGCCATCAAGAAGAGAGCAGACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 913

Db

3581 TCGCCATCAAGAAGAGAGCAGACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 3640

Qy

914 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCGCACCCCGCGCGCTGA 973

Db

3641 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCGCACCCCGCGCGCTGA 3700

Qy

974 AGAAGAGAGAGCGTGACCGTCTGTGACGCTGGCGGACGCTACTTCAAGGTGCGCCCTGG 1033

Db

3701 AGAAGAGAGAGCGTGACCGTCTGTGACGCTGGCGGACGCTACTTCAAGGTGCGCCCTGG 3760

Qy

1034 ACAGAGACTTCCGCAAGTACACCGCCCTTCCATCCCGCAGCATCAACAGAGACCCCGCG 1093

Db

3761 ACAGAGACTTCCGCAAGTACACCGCCCTTCCATCCCGCAGCATCAACAGAGACCCCGCG 3820

Qy

1094 GCATCCGCTACCAAGTACACCGTGTGCCCCAGGGCTGGAGGAGCGAGCCCGAGCATCTTCC 1153

Db

3821 GCATCCGCTACCAAGTACACCGTGTGCCCCAGGGCTGGAGGAGCGAGCCCGAGCATCTTCC 3880

Qy

1154 AGAGCAGCATGACCAAGATCCTTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 1213

Db

3881 AGAGCAGCATGACCAAGATCCTTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 3940

Qy

1214 ACCAGGCCCCCTGTGTGTGAGCGAGCGCATCTGGAGATCGGCGCAGCATCGCGCGCGCG 1273

Db

3941 ACCAGGCCCCCTGTGTGTGAGCGAGCGCATCTGGAGATCGGCGCAGCATCGCGCGCGCG 4000

Qy

1274 AGGAGTGGCGCAAGCAGCTGTGTGCGCTGGGGCTTCAACACCCCGCGCAAGAGCAGCA 1333



Db 4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAACACCCCGACAAAGACACCAGA 4060  
Qy 1334 AGGAGCCCCCTTCTTCCCATCGAGCTGCACCCCGACAAAGTGGACCGTGAGCCCATCG 1393  
Db 4061 AGGAGCCCCCTTCTTCCCATCGAGCTGCACCCCGACAAAGTGGACCGTGAGCCCATCG 4120  
Qy 1394 AGCTGCCGAGAGAGAGAGCTGGACCGTGAAGCAGATCCAGAAAGCTGGTGGCAAGCTGA 1453  
Db 4121 AGCTGCCGAGAGAGAGAGCTGGACCGTGAAGCAGATCCAGAAAGCTGGTGGCAAGCTGA 4180  
Qy 1454 ACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCAGCTGTGCAAGCTGTGCGCG 1513  
Db 4181 ACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCAGCTGTGCAAGCTGTGCGCG 4240  
Qy 1514 GGGCCAAAGCCCTGACCGACATCTGTGCCCTGACCGAGAGGCCGAGCTGGAGCTGGCG 1573  
Db 4241 GGGCCAAAGCCCTGACCGACATCTGTGCCCTGACCGAGAGGCCGAGCTGGAGCTGGCG 4300  
Qy 1574 AGAACCGGAGATCTTGGCGGAGCCGTGACGGGTGTACTACGACCCCAAGCAAGGACC 1633  
Db 4301 AGAACCGGAGATCTTGGCGGAGCCGTGACGGGTGTACTACGACCCCAAGCAAGGACC 4360  
Qy 1634 TGGTGGCCGAGATCCAGAGCAGGSCCAGCACAGTGGACCTTACAGATCTTACAGGAGC 1693  
Db 4361 TGGTGGCCGAGATCCAGAGCAGGSCCAGCACAGTGGACCTTACAGATCTTACAGGAGC 4420  
Qy 1694 CTTTCAAGAACCTGAAGACCCGCAAGTAGCCCAAGATGGGACCCGCCACACCAAGCAGG 1753  
Db 4421 CTTTCAAGAACCTGAAGACCCGCAAGTAGCCCAAGATGGGACCCGCCACACCAAGCAGG 4480  
Qy 1754 TGAAGCAGCTGACCGAGGCGGTGCAAGAGATGCCATGAGAGCATCTGTATCTGGGCA 1813  
Db 4481 TGAAGCAGCTGACCGAGGCGGTGCAAGAGATGCCATGAGAGCATCTGTATCTGGGCA 4540  
Qy 1814 AGACCCCAAGTTCCGCTGCGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGACCGACT 1873  
Db 4541 AGACCCCAAGTTCCGCTGCGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGACCGACT 4600  
Qy 1874 ACTGGAGGCCACCTGATCCCGAGTGGAGTTGTTGAACACCCCGCCCTGGTGAAGC 1933  
Db 4601 ACTGGAGGCCACCTGATCCCGAGTGGAGTTGTTGAACACCCCGCCCTGGTGAAGC 4660  
Qy 1934 TGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGACGGCG 1993  
Db 4661 TGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGACGGCG 4720  
Qy 1994 CGCCCAACCGGAGACCAAGATCGGCAAGCGCGCTTACGTACCGACCGGGCCGGGCA 2053  
Db 4721 CGCCCAACCGGAGACCAAGATCGGCAAGCGCGCTTACGTACCGACCGGGCCGGGCA 4780  
Qy 2054 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGTGCAGGCGCATCCAGCTGG 2113  
Db 4781 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGTGCAGGCGCATCCAGCTGG 4840  
Qy 2114 CCTCGAGGACCGGCGAGGAGTGAACATCTGTACCGACAGCAGTACGCCCTGGGCA 2173  
Db 4841 CCTCGAGGACCGGCGAGGAGTGAACATCTGTACCGACAGCAGTACGCCCTGGGCA 4900  
Qy 2174 TCATCCAGGCCCGGCGGACCAAGAGCGAGGAGCTGGTGAACAGATCATCGAGCAGC 2233  
Db 4901 TCATCCAGGCCCGGCGGACCAAGAGCGAGGAGCTGGTGAACAGATCATCGAGCAGC 4960  
Qy 2234 TGATCAAGAGAGAGGTGTACCTGAGCTGGTGGTCCGCCCAAGGGCATCCGGGCA 2293  
Db 4961 TGATCAAGAGAGAGGTGTACCTGAGCTGGTGGTCCGCCCAAGGGCATCCGGGCA 5020  
Qy 2294 AGGAGCAGATCGACAAGCTGGTGAAGGGCATCCGCAAGGTGTCTTCTGGACGGCA 2353  
Db 5021 AGGAGCAGATCGACAAGCTGGTGAAGGGCATCCGCAAGGTGTCTTCTGGACGGCA 5080  
Qy 2354 TCGATGGCGGATCTGTGATCTACAGTACATGAGCAGCTGTGAGGCGAGCGCGCC 2413  
Db 5081 TCGATGGCGGATCTGTGATCTACAGTACATGAGCAGCTGTGAGGCGAGCGCGCC 5140

Qy 2414 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 5141 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 5178

RESULT 12  
ACA03547  
ID ACA03547 standard; DNA; 2457 BP.

XX ACA03547;  
XX AC  
XX DT 22-MAY-2003 (first entry)

XX Synthetic DNA ecoding immunogenic HIV peptide #30.

XX Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX Synthetic.

XX WO2003004657-A1.  
XX PD 16-JAN-2003.

XX PF 05-JUL-2002; 2002WO-US021421.

XX PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349728P.  
PR 16-JAN-2002; 2002US-0349793P.  
PR 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX Zur Megede J, Barnett SW, Lian Y;

XX WPI; 2003-221602/21.

XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
PT especially humans.

XX Example 1; Fig 35; 262pp; English.

XX The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide

SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.8%; Score 2428.6; DB 8; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 3.2e-292;  
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 1 GTCGACGCCACCATGCGCCGAGCCATGAGCCAGGCCAGCCCAACATCTGTATGCAG 60

Db 1 GTCGACGCCACCATGCGCCGAGCCATGAGCCAGGCCAGCCCAACATCTGTATGCAG 60

Qy 61 CGCAGCAACTTCAAGGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGC 120

Db 61 CGCAGCAACTTCAAGGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGC 120

Qy	121	CACATCGCCCGCAACTGCGCGCCCGCCCGCAAGAAGGGCTGCTGGAAGTGGCGCAAGGAG	180
Db	121	CACATCGCCCGCAACTGCGCGCCCGCCCGCCGAAGAAGGGCTGCTGGAAGTGGCGCAAGGAG	180
Qy	181	GGCCACCAAGATGAAGGACTGACACGAGGCGCAGGCCCAACTTCTTCCGCGAGGACTGCGCC	240
Db	181	GGCCACCAAGATGAAGGACTGACACGAGGCGCAGGCCCAACTTCTTCCGCGAGGACTGCGCC	240
Qy	241	TTCCCCCAAGGCGAAGGCCCGCGAGTTCCCAAGCGAGCAAGACCGCGCCAAACAGGCCCAAC	300
Db	241	TTCCCCCAAGGCGAAGGCCCGCGAGTTCCCAAGCGAGCAAGACCGCGCCAAACAGGCCCAAC	300
Qy	301	AGCCGCGAGCTGCAAGTGCAGGCGGCGCAAAACCCCGCAGCGAGGCGCGCGAGCGCCAG	360
Db	301	AGCCGCGAGCTGCAAGTGCAGGCGGCGCAAAACCCCGCAGCGAGGCGCGCGAGCGCCAG	360
Qy	361	GGCACCTTGAACTTCCCCCAGATCAACCTGTGCGAGCGCCCTTGTTGAGCATCAAGGTG	420
Db	361	GGCACCTTGAACTTCCCCCAGATCAACCTGTGCGAGCGCCCTTGTTGAGCATCAAGGTG	420
Qy	421	GGCGGCGAGATCAAGGAGGCCCTGCTGGAACAACCGCGCGCAGCAACCGTCTCGAGGAG	480
Db	421	GGCGGCGAGATCAAGGAGGCCCTGCTGCGCAACCGCGCGCAGCAACCGTCTCGAGGAG	480
Qy	481	ATGAGCCTGCCCGCAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAACAG	540
Db	481	ATGAGCCTGCCCGCAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAACAG	540
Qy	541	GTGCGCCAGTAGCACCAAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAACCGTG	600
Db	541	GTGCGCCAGTAGCACCAAGATCCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCAACCGTG	600
Qy	601	CTGATCGGCGCCCAACCCCGTGAAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Db	601	CTGATCGGCGCCCAACCCCGTGAAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Qy	661	ACCTGGAATTCCTCCATCAGCCCATCGAGACCTGCCCGTGAAGCTGAGCCCGGCATG	720
Db	661	ACCTGGAATTCCTCCATCAGCCCATCGAGACCTGCCCGTGAAGCTGAGCCCGGCATG	720
Qy	721	GACGGCCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAAGATCAAGGCCCTCACCGCC	780
Db	721	GACGGCCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAAGATCAAGGCCCTCACCGCC	780
Qy	781	ATCTCGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGTCGCGCCCGCAGAAACCCCTAC	840
Db	781	ATCTCGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGTCGCGCCCGCAGAAACCCCTAC	840
Qy	841	AACACCCCGTTCGCGCATCAAGAAGAAGACAGCACCAAGTCGCGCCCGCAGAAACCCCTAC	900
Db	841	AACACCCCGTTCGCGCATCAAGAAGAAGACAGCACCAAGTCGCGCCCGCAGAAACCCCTAC	900
Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTCGGAGGTGCACTGGGCAATCCCCCAC	960
Db	901	TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTCGGAGGTGCACTGGGCAATCCCCCAC	960
Qy	961	CCCGCGGCGCTGAAGAAGAAGAGCGTGACCGTGTGAGCTGGGCGACGCCCTACTTC	1020
Db	961	CCCGCGGCGCTGAAGAAGAAGAGCGTGACCGTGTGAGCTGGGCGACGCCCTACTTC	1020
Qy	1021	AGCGTGCCCTGGACGAGGACTTCGCGAAGTACACCGCTTCAACATCCCGAGCATCAAC	1080
Db	1021	AGCGTGCCCTGGACGAGGACTTCGCGAAGTACACCGCTTCAACATCCCGAGCATCAAC	1080
Qy	1081	AACGAGACCCCGGCATCCGTACAGTACAAACGTGTGTCGCCCGCAGGGCTGGAAGGGCAGC	1140
Db	1081	AACGAGACCCCGGCATCCGTACAGTACAAACGTGTGTCGCCCGCAGGGCTGGAAGGGCAGC	1140
Qy	1141	CCGAGCATCTTCCAGAGCAGATGACCAAGATCTCGAGCCCTTTCGCGCCCGCAACCCC	1200
Db	1141	CCGAGCATCTTCCAGAGCAGATGACCAAGATCTCGAGCCCTTTCGCGCCCGCAACCCC	1200

QY	1201	GAGATCGTGATCTTACAGAGGCCCCCTCTGTACGTGGGCAGCGCACTCTGGAGATCGCGCAGCAC	1260
DB	1201	GAGATCGTGATCTTACAGAGGCCCCCTCTGTACGTGGGCAGCGCACTCTGGAGATCGCGCAGCAC	1260
QY	1261	CGCGCCAAGATTCGAGGAGCTGCGCAAGCACCTGCTGGCTGGGGCTTCCACACCCCGCAG	1320
DB	1261	CGCGCCAAGATTCGAGGAGCTGCGCAAGCACCTGCTGGCTGGGGCTTCCACACCCCGCAG	1320
QY	1321	AAGAAGCACACAGAAAGGAGCCCCCTTCTTGCCCAT-----CGAGTGCACACCCGCAAG	1374
DB	1321	AAGAAGCACACAGAAAGGAGCCCCCTTCTTGTGTGGCTATCAGAGTGCACCCCGCAAG	1380
QY	1375	TGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGATCCAG	1434
DB	1381	TGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGATCCAG	1440
QY	1435	AAGCTGCTGGCAGACTGAACTGGCGCAGCCAGATCTTACCCGGCATCAAGGTGCCCGAG	1494
DB	1441	AAGCTGCTGGCAGACTGAACTGGCGCAGCCAGATCTTACCCGGCATCAAGGTGCCCGAG	1500
QY	1495	CTGTGCAAGCTGTGCGCGCGCAAGGCCCTTGACCGACATCGTGCCTCCCTGACCCGAGGAG	1554
DB	1501	CTGTGCAAGCTGTGCGCGCGCAAGGCCCTTGACCGACATCGTGCCTCCCTGACCCGAGGAG	1560
QY	1555	GCGAGCTGAGCTGGCGCGAGAACCGCGGAGATCTTGGCGGAGCCGTGCAACGCGGTGTAC	1614
DB	1561	GCGAGCTGAGCTGGCGCGAGAACCGCGGAGATCTTGGCGGAGCCGTGCAACGCGGTGTAC	1620
QY	1615	TACGACCCACAGCAAGNACTGTGTGGCGGAGATCCAGNACGAGGCCACGACAGTGCAGC	1674
DB	1621	TACGACCCACAGCAAGNACTGTGTGGCGGAGATCCAGNACGAGGCCACGACAGTGCAGC	1680
QY	1675	TACCAGATCTACCAGGAGCCCTTCAAGAACCTTGAAGACCGGCAGAGTACGCAAGATGCGC	1734
DB	1681	TACCAGATCTACCAGGAGCCCTTCAAGAACCTTGAAGACCGGCAGAGTACGCAAGATGCGC	1740
QY	1735	ACGCGCCACACCAACGACGTGAAGCAGCTGACGAGGCCGTGCAAGAGTTCGCCATGCGAG	1794
DB	1741	ACGCGCCACACCAACGACGTGAAGCAGCTGACGAGGCCGTGCAAGAGTTCGCCATGCGAG	1800
QY	1795	AGCATCTGTGATCTGGGGCAGAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTG	1854
DB	1801	AGCATCTGTGATCTGGGGCAGAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTG	1860
QY	1855	GAGACCTGTGTGGACCCGACTACTGGCAGGCCACCTGGATCCCCAGTGGGAGTTCGTGAAC	1914
DB	1861	GAGACCTGTGTGGACCCGACTACTGGCAGGCCACCTGGATCCCCAGTGGGAGTTCGTGAAC	1920
QY	1915	ACCCCCCTCTGTGTGAAGTGTGGTACCACTGAGTGGAGAGGCCCATCATTCGCGCCGAG	1974
DB	1921	ACCCCCCTCTGTGTGAAGTGTGGTACCACTGAGTGGAGAGGCCCATCATTCGCGCCGAG	1980
QY	1975	ACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGGCTACG	2034
DB	1981	ACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGGCTACG	2040
QY	2035	ACGACCGGGCGCGCAGAGATCTGTAGCTGCAACGAGACCAACCAACAGAGACCGAG	2094
DB	2041	ACGACCGGGCGCGCAGAGATCTGTAGCTGCAACGAGNCAACCAACAGAGACCGAG	2100
QY	2095	CTGAGGCCATCAGACTGGCCCTCTGAGGACAGCGGAGCGAGGTGAACATCTGTGACCGAG	2154
DB	2101	CTGAGGCCATCAGACTGGCCCTCTGAGGACAGCGGAGCGAGGTGAACATCTGTGACCGAG	2160
QY	2155	AGCCAGTACGCCCTTGGGCATCATCTAGGCCACGCCCGACAGAGCGAGCGAGCTGGTG	2214
DB	2161	AGCCAGTACGCCCTTGGGCATCATCTAGGCCACGCCCGACAGAGCGAGCGAGCTGGTG	2220
QY	2215	AACCAGATCATCGAGCAGCTGTACAAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCC	2274
DB	2221	AACCAGATCATCGAGCAGCTGTACAAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCC	2280
QY	2275	CACAAGGGCATCTGGCGGCAACGAGCAGATCGACAAGCTGTGAGCAGAGGCGCATCCGCAAG	2334

Db 2281 CACAAGGGCATCGGGGCAACAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAG 2340  
QY 2335 GTGCTGTTCTCTGAGACGGCATCGATGCGGGCATCGTGAATCTACAGTACATGAGACACCTG 2394  
Db 2341 GTGCTGTTCTCTGAGACGGCATCGATGCGGGCATCGTGAATCTACAGTACATGAGACACCTG 2400  
QY 2395 TACGTGGGAGCGGGGCGCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCCGT 2451  
Db 2401 TACGTGGGAGCGGGGCGCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCCGT 2457

RESULT 13  
ID ADC13265  
ID ADC13265 standard; DNA; 2457 BP.  
XX  
AC ADC13265;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA of HIV construct p2Pol-opt-YM\_C SEQ ID NO 44.  
XX  
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX  
OS Human immunodeficiency virus.  
XX  
PN WO2003004620-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 05-JUL-2002; 2002WO-US021420.  
XX  
PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX WPI; 2003-221593/21.  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.

XX Disclosure; Fig 41; 301pp; English.  
XX  
CC The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramuscularly, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.

SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;  
Query Match 98.8%; Score 2428.6; DB 10; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 3.2e-292;  
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 1 GTGAGCGCCACCATGGCGGCGCATGAGCCAGGCCACAGGCCCAATCTGTATGAG 60  
Db 1 GTGAGCGCCACCATGGCGGCGCATGAGCCAGGCCACAGGCCCAATCTGTATGAG 60

QY 61 CGCAGCAACTTCAAGGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGTGGCGCAAGAGGGC 120  
Db 61 CGCAGCAACTTCAAGGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGTGGCGCAAGAGGGC 120  
QY 121 CACATCGCCCCCAACTGCGCGCCCCCGCAGAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCCCCAACTGCGCGCCCCCGCAGAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
QY 181 GGCACACAGATGAAGGACTGCACCGAGCGCCAGAGCAATCTTCTCCGCGAGGACCTGGCC 240  
Db 181 GGCACACAGATGAAGGACTGCACCGAGCGCCAGAGCAATCTTCTCCGCGAGGACCTGGCC 240  
QY 241 TTCCGCCAGGCAAGGCCCGCGAGTTCCCGCAGCAGAGCAACCGCGCCCAACAGAGCCACCC 300  
Db 241 TTCCGCCAGGCAAGGCCCGCGAGTTCCCGCAGCAGAGCAACCGCGCCCAACAGAGCCACCC 300  
QY 301 AGCGCGAGCTGCAGGTGCGGGGACAAACCCCGCAGCGAGGGCGCGCGCCAGGCCAG 360  
Db 301 AGCGCGAGCTGCAGGTGCGGGGACAAACCCCGCAGCGAGGGCGCGCGCCAGGCCAG 360  
QY 361 GGCACCTGAACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGCATCAAGTGTG 420  
Db 361 GGCACCTGAACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGCATCAAGTGTG 420  
QY 421 GCGGCGCAGATCAAGGAGGCCCTGCTGGAACACCGCGCGCGAGCACACCGTGTGTGAGGAG 480  
Db 421 GCGGCGCAGATCAAGGAGGCCCTGCTGGAACACCGCGCGCGAGCACACCGTGTGTGAGGAG 480  
QY 481 ATGAGCTGTGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCAACAAG 540  
Db 481 ATGAGCTGTGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCAACAAG 540  
QY 541 GTGCGCGATGATGACACAGATCTGTGATCGAGATCTGCGGCGAAGAGGCCATCGGCAACCTG 600  
Db 541 GTGCGCGATGATGACACAGATCTGTGATCGAGATCTGCGGCGAAGAGGCCATCGGCAACCTG 600  
QY 601 CTGATCGGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAACTTCCCCCATCAGCCCCCATCGAGACCGGTGCGCGCTGAAGCTGAAAGCCCGCATG 720  
Db 661 ACCCTGAACTTCCCCCATCAGCCCCCATCGAGACCGGTGCGCGCTGAAGCTGAAAGCCCGCATG 720  
QY 721 GACGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCCGCC 780  
Db 721 GACGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCCGCC 780  
QY 781 ATCTGCGAGGAGATGGAGAAGGGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGGAGAAGGGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTAC 840  
QY 841 AACACCCCGGTGTTGCCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGGTGTTGCCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
QY 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCAATCCGCCAC 960  
Db 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCAATCCGCCAC 960  
QY 961 CCCGCGGCGCTGAAGAAGAGAGAGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTACTTC 1020  
Db 961 CCCGCGGCGCTGAAGAAGAGAGAGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTACTTC 1020  
QY 1021 AGCGTGGCCCCCTGGAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGGCCCCCTGGAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
QY 1081 AACGAGACCCCGCGCATCCGCTACACAGTACACAGTGTGCCCCAGGGGTGGAGGGGAGC 1140  
Db 1081 AACGAGACCCCGCGCATCCGCTACACAGTACACAGTGTGCCCCAGGGGTGGAGGGGAGC 1140  
QY 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGTGGAGCGCTTCCGCGCGCCGCAACCCC 1200

```
Db 1141 |CCAGCATCTTCCAGAGCAGATGACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCCC| 1200
Qy 1201 |GAGATCGTGTATCTACCAAGGCCCCCTCTAGCTGGGCGAGCACTTGAGATCGGCCAGCAC| 1260
Db 1201 |GAGATCGTGTATCTACCAAGGCCCCCTCTAGCTGGGCGAGCACTTGAGATCGGCCAGCAC| 1260
Qy 1261 |CGCGCAAGATCGAGAGCTGGCGAAGCACTGCTGCGCTGGGGCTTCAACAACCCCGAC| 1320
Db 1261 |CGCGCAAGATCGAGAGCTGGCGAAGCACTGCTGCGCTGGGGCTTCAACAACCCCGAC| 1320
Qy 1321 |AAGAGCACCAGAGAGGCCCCCTTCTGCGCAT-----CGAGCTGCAACCCCGACAAG| 1374
Db 1321 |AAGAGCACCAGAGAGGCCCCCTTCTGCGCAT-----CGAGCTGCAACCCCGACAAG| 1380
Qy 1375 |TGACACCTGCGAGCTGCGCGGCGCCAGAGAGCTGGACCGTGAACGACATCCAG| 1434
Db 1381 |TGNACCGTGCAGCCCTCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAG| 1440
Qy 1435 |AAGCTGCTGGGCAAGCTGAATCTGGGCGAGCCAGCAATCTACCCCGGCATCAAGGTCGCGCAG| 1494
Db 1441 |AAGCTGCTGGGCAAGCTGAATCTGGGCGAGCCAGATCTTACCCCGGCATCAAGGTCGCGCAG| 1500
Qy 1495 |CTGTGCAAGCTGCTGCGCGGCGCCAGGCCCCCTGACCGACATCTGCCCCCTGACCGAGGAG| 1554
Db 1501 |CTGTGCAAGCTGCTGCGCGGCGCCAGGCCCCCTGACCGACATCTGCCCCCTGACCGAGGAG| 1560
Qy 1555 |GCCGAGCTGAGCTGGCGCGAGAACCGGAGATCTGCGCGAGCCCGTGCACGCGCTGTAC| 1614
Db 1561 |GCCGAGCTGAGCTGGCGCGAGAACCGGAGATCTGCGCGAGCCCGTGCACGCGCTGTAC| 1620
Qy 1615 |TAGCACCCCGAGAGGACCTGGTGGCGAGATCCAGAGCAGGCGCCACGACCAAGTGGACC| 1674
Db 1621 |TAGCACCCCGAGAGGACCTGGTGGCGAGATCCAGAGCAGGCGCCACGACCAAGTGGACC| 1680
Qy 1675 |TACGAGATCTACGAGAGCCCTTCAGAACTGAAGACCGGCAAGTACGCCAAGATGCGC| 1734
Db 1681 |TACGAGATCTACGAGAGCCCTTCAGAACTGAAGACCGGCAAGTACGCCAAGATGCGC| 1740
Qy 1735 |ACCGCCACACCAAGACGCTGAAGACGCTGACGAGGCGCTGAGAGATCGCCATGGAG| 1794
Db 1741 |ACCGCCACACCAAGACGCTGAAGACGCTGACGAGGCGCTGAGAGATCGCCATGGAG| 1800
Qy 1795 |AGCATCGTGTATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGAGGACCTGG| 1854
Db 1801 |AGCATCGTGTATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGAGGACCTGG| 1860
Qy 1855 |GAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAC| 1914
Db 1861 |GAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAC| 1920
Qy 1915 |ACCCGCCCTGCTGAAGCTGTGTACAGCTGAGAGAGGAGCCCATCATCGGCGCGGAG| 1974
Db 1921 |ACCCGCCCTGCTGAAGCTGTGTACAGCTGAGAGAGGAGCCCATCATCGGCGCGGAG| 1980
Qy 1975 |ACCTTCTACGTGAGCGCGCGCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTG| 2034
Db 1981 |ACCTTCTACGTGAGCGCGCGCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTG| 2040
Qy 2035 |ACCGACGGGGCGCGAAGATGCTGAGCTGACCGAGACCAACCAACCAAGAGACCGAG| 2094
Db 2041 |ACCGACGGGGCGCGAAGATGCTGAGCTGACCGAGACCAACCAACCAAGAGACCGAG| 2100
Qy 2095 |CTGACGSCCATCCAGCTGGCCCTGCGAGGACAGCGGAGGCTGACATCTGTCGCGAC| 2154
Db 2101 |CTGACGSCCATCCAGCTGGCCCTGCGAGGACAGCGGAGGCTGACATCTGTCGCGAC| 2160
Qy 2155 |AGCCAGTACCCCTGGGCAATCATCCAGGCGCCAGCCGACAAAGAGCGAGAGCTGGTG| 2214
Db 2161 |AGCCAGTACCCCTGGGCAATCATCCAGGCGCCAGCCGACAAAGAGCGAGAGCTGGTG| 2220
Qy 2215 |AACCAGATCATCAGACGCTGATCAAGAGAGAGAGTGTACTGTGAGCTGGTGCCCGCC| 2274
```

```
Db 2221 |AACGAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGTCCCGCC| 2280
Qy 2275 |CACAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAAGCAAGGGCATCCGCAAG| 2334
Db 2281 |CACAGGGCATCGCGGCAACGAGCAGATCGACAGCTGGTGAAGCAAGGGCATCCGCAAG| 2340
Qy 2335 |GTCTGTCTCTGACGCGCATCGATGCGGCGCATCGTGTATCTACAGTACATGAGACGCTG| 2394
Db 2341 |GTCTGTCTCTGACGCGCATCGATGCGGCGCATCGTGTATCTACAGTACATGAGACGCTG| 2400
Qy 2395 |TACGTGGGCGAGCGCGCCCTTAGATCGATTAAGCTTCCCGGGCTAGACCGGT| 2451
Db 2401 |TACGTGGGCGAGCGCGCCCTTAGATCGATTAAGCTTCCCGGGCTAGACCGGT| 2457

RESULT 14
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein; immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef; immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
XX
SY Synthetic.
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR ) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
WPI; 2002-154920/20.
XX
New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.
XX
Claim 1; Fig 8; 233pp; English.
XX
The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef (I). (I) have immunostimulant activity and can be used in gene therapy. The HIV type C polynucleotides are useful in applications including DNA immunisation, generation of packaging cell lines, and production of HIV type C proteins. The polynucleotides are particularly useful in gene therapy and DNA immunisation applications. ABL39942 to ABL40054 and CC ABB06204 to ABB06215 represent sequences used in the exemplification of the present invention
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;
Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
Best Local Similarity 99.3%; Pred. No. 1.4e-290;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;
Qy 1 GTGCGACCCACCATGCGCGAGGCCATGAGCCAGGCCAGGCCAACATCTCTGATGCGAG 60
```

Db 1 GTCAGCCACCATGCGCGAGCCCATGAGCCAGGCCACAGCGCAACATCCTGATGAG 60  
Qy 61 CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGGCAAGAGGGC 120  
Db 61 CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGGCAAGAGGGC 120  
Qy 121 CACATCGCCCGCAACTGCGCGCCCCCGCGAAGAGGGCTGCTGGAAGTGGCGAAGGAG 180  
Db 121 CACATCGCCCGCAACTGCGCGCCCCCGCGAAGAGGGCTGCTGGAAGTGGCGAAGGAG 180  
Qy 181 GGCCACAGATGAAGAGCTGACCGAGCGCCAGCGCCAACTTCTTTCGCGAGGACCTGGCC 240  
Db 181 GGCCACAGATGAAGAGCTGACCGAGCGCCAGCGCCAACTTCTTTCGCGAGGACCTGGCC 240  
Qy 241 TTCCCGCAGGCGAGGCCCGGAGTCCCCAGCGAGCAGAACCGCGCCAAAGAGGAGGAGG 300  
Db 241 TTCCCGCAGGCGAGGCCCGGAGTCCCCAGCGAGCAGAACCGCGCCAAAGAGGAGGAGG 300  
Qy 301 AGCGCGAGCTGCAAGGTGCGCGCGCAAAACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360  
Db 301 AGCGCGAGCTGCAAGGTGCGCGCGCAAAACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360  
Qy 361 GGCACTTGAACTTCCCGCCAGATCACCTGTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG 420  
Db 361 GGCACTTGAACTTCCCGCCAGATCACCTGTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG 420  
Qy 421 GGCGGCAGATCAAGGAGGCCCTGTGACACCGCGCGCAGCGAGCGCGCGCGAGCGAGGAG 480  
Db 421 GGCGGCAGATCAAGGAGGCCCTGTGACACCGCGCGCAGCGAGCGCGCGCGAGCGAGGAG 480  
Qy 481 ATGAGCCTTGCCTCGGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGCGCTTTCATCAAG 540  
Db 481 ATGAGCCTTGCCTCGGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGCGCTTTCATCAAG 540  
Qy 541 GTGCGCGATGACACCGATCCTGATCGAGATCTGCGCGAAGAGGCCATCGGCACCGTG 600  
Db 541 GTGCGCGATGACACCGATCCTGATCGAGATCTGCGCGAAGAGGCCATCGGCACCGTG 600  
Qy 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCGAGCTGGGCTGC 660  
Db 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCGAGCTGGGCTGC 660  
Qy 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG 720  
Db 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG 720  
Qy 721 GACGCGCCAAAGGTGAAGCTGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCC 780  
Db 721 GACGCGCCAAAGGTGAAGCTGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCC 780  
Qy 781 ATCTCGGAGGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCGAGAACCCCTAC 840  
Db 781 ATCTCGGAGGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCGAGAACCCCTAC 840  
Qy 841 AACACCCCGTGTTCGCGCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGTGTTCGCGCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Qy 901 TTCCGCGAGCTGAACAAAGCGCACCCAGACTTCTGGGAGGTGACGTGGGCAATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAAGCGCACCCAGACTTCTGGGAGGTGACGTGGGCAATCCCCAC 960  
Qy 961 CCGCGCGCTGAGAGAGAGAGCGTGAACCGTGTGGAGCTGGGCGAGCGCTACTTC 1020  
Db 961 CCGCGCGCTGAGAGAGAGAGCGTGAACCGTGTGGAGCTGGGCGAGCGCTACTTC 1020  
Qy 1021 AGCGTGGCCCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
Db 1021 AGCGTGGCCCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
Qy 1081 AACGAGACCCCGGATCCCGTACCAAGTACAACTGTGTGCGCCAGGCTGGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGGATCCCGTACCAAGTACAACTGTGTGCGCCAGGCTGGAAGGGCAGC 1140

Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTTCGCGCGCCGCAACCCC 1200  
Db 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTTCGCGCGCCGCAACCCC 1200  
Qy 1201 GAGATCGTGTATCTACCA-----GGCCCCCTGTAGCTGGGCGAGCACTTGGAGATCGGC 1254  
Db 1201 GAGATCGTGTATCTACCAAGTACATGGACGACCTGTAGCTGGGCGAGCACTTGGAGATCGGC 1260  
Qy 1255 CAGCACCGCGCAAGATCGAGGAGCTGCGAAGCACTGTCTGCGCTGGGCTTTCACACCC 1314  
Db 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGAAGCACTGTCTGCGCTGGGCTTTCACACCC 1320  
Qy 1315 CCGACAAGAAGACCAAGAGGAGGCCCTTCTCTGCCAT-----CGAGCTGCACCCC 1368  
Db 1321 CCGACAAGAAGACCAAGAGGAGGCCCTTCTCTGGAATGGGCTAGAGCTGCACCCC 1380  
Qy 1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1428  
Db 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1440  
Qy 1429 ATCCAGAAGCTGTGTGGCAAGCTGAATGGGCGACCGAGATCTTACCCCGGATCAAGGTG 1488  
Db 1441 ATCCAGAAGCTGTGTGGCAAGCTGAATGGGCGACCGAGATCTTACCCCGGATCAAGGTG 1500  
Qy 1489 CGCCAGCTGTCAAGCTGTGCGCGGCCAAGSCCTGACCGACATCTGTGCCCTGACC 1548  
Db 1501 CGCCAGCTGTCAAGCTGTGCGCGGCCAAGSCCTGACCGACATCTGTGCCCTGACC 1560  
Qy 1549 GAGGAGCCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGC 1608  
Db 1561 GAGGAGCCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGC 1620  
Qy 1609 GTGTACTACGACCCCGAAGAGACCTGTGTGGCGAGATCCAGAAAGAGGCGCACGACAG 1668  
Db 1621 GTGTACTACGACCCCGAAGAGACCTGTGTGGCGAGATCCAGAAAGAGGCGCACGACAG 1680  
Qy 1669 TGAACCTTACCAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTAGTGCACAG 1728  
Db 1681 TGAACCTTACCAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTAGTGCACAG 1740  
Qy 1729 ATGCGCACCGCCCAACCAACGAGTGAAGCAGTGAACGAGGCGGTGCAAGAGATCGCC 1788  
Db 1741 ATGCGCACCGCCCAACCAACGAGTGAAGCAGTGAACGAGGCGGTGCAAGAGATCGCC 1800  
Qy 1789 ATGAGAGCATGTGTATCTGGGCGAGACCCCAAGTTCGCGCTGCCATCCAGAAAGAG 1848  
Db 1801 ATGAGAGCATGTGTATCTGGGCGAGACCCCAAGTTCGCGCTGCCATCCAGAAAGAG 1860  
Qy 1849 ACCTGGAGACCTGTGTGGACCGACTCTGTGAGCGCACCTGGATCCCGGAGTGGAGTTC 1908  
Db 1861 ACCTGGAGACCTGTGTGGACCGACTCTGTGAGCGCACCTGGATCCCGGAGTGGAGTTC 1920  
Qy 1909 GTGAACACCCCGCTGTGTGAAGCTGTGTGTACACAGCTGGAGAGGAGCCCATCATCGGC 1968  
Db 1921 GTGAACACCCCGCTGTGTGAAGCTGTGTGTACACAGCTGGAGAGGAGCCCATCATCGGC 1980  
Qy 1969 GCGAGACCTTCTACGTGGAAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCG 2028  
Db 1981 GCGAGACCTTCTACGTGGAAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCG 2040  
Qy 2029 TAGGTGACCGACCGGCGCGGAGAGATGTGAGCTGTGAGCTGTGAGACCAACCAAGAG 2088  
Db 2041 TAGGTGACCGACCGGCGCGGAGAGATGTGAGCTGTGAGCTGTGAGACCAACCAAGAG 2100  
Qy 2089 ACCGAGTGCAGGCACTTCCAGCTGGCCCTGACGACAGCGCGCAGCGAGGTGAACATCGTG 2148  
Db 2101 ACCGAGTGCAGGCACTTCCAGCTGGCCCTGACGACAGCGCGCAGCGAGGTGAACATCGTG 2160  
Qy 2149 ACCGAGTGCAGGTACGCCCTTGGGCAATCATCAGGCCCGACCCGACAGAGCGAGGAGCGAG 2208  
Db 2161 ACCGAGTGCAGGTACGCCCTTGGGCAATCATCAGGCCCGACCCGACAGAGCGAGGAGCGAG 2220

QY 2209 CTGGTGAACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTG 2268  
Db 2221 CTGGTGAACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTG 2280  
QY 2269 CCCGCCCAAGGGCATCGCGGCGAAGCAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2328  
Db 2281 CCCGCCCAAGGGCATCGCGGCGAAGCAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340  
QY 2329 CGCAAGGTGCTTCTTGGACCGGCATCGATCGCGGCATCGTGATCTACCACTACATGGAC 2388  
Db 2341 CGCAAGGTGCTTCTTGGACCGGCATCGATCGCGGCATCGTGATCTACCACTACATGGAC 2400  
QY 2389 GACCTGTACGTGGCAGCGCGGCGCTAGGATCGATTAAGAGCTTCCGGGGCTAGCAC 2448  
Db 2401 GACCTGTACGTGGCAGCGCGGCGCTAGGATCGATTAAGAGCTTCCGGGGCTAGCAC 2460  
QY 2449 GGTGAATTC 2457  
Db 2461 GGTGAATTC 2469  
RESULT 15  
ID ADM73764 standard; DNA; 2469 BP.  
AC ADM73764;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE HIV-1 polynucleotide #7.  
XX  
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
KW HIV type C protein; immunostimulant.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN US2003223961-A1.  
XX  
PD 04-DEC-2003.  
XX  
PP 05-JUL-2001; 2001US-00899575.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
XX  
PA (MEGE/) MEGEDE J Z.  
PA (BARN/) BARNETT S W.  
PA (ENGE/) ENGELBRECHT S.  
PA (RENS/) RENSBERG E J V.  
XX  
PI Megede JZ., Barnett SW, Engelbrecht S, Rensburg EJV;  
XX WPI; 2004-060515/06.  
XX  
DR New expression cassette comprising a polynucleotide sequence encoding an  
PT HIV pol polypeptide, useful in eliciting an immune response, in DNA  
PT immunization, generating of packaging cell lines or in producing HIV Type  
PT C proteins.  
XX  
PS Claim 1; SEQ ID NO 30; 160pp; English.  
XX  
CC The invention relates to an expression cassette comprising a  
CC polynucleotide sequence encoding an HIV pol polypeptide. The invention  
CC also relates to a recombinant expression system for use in a host cell  
CC comprising an expression cassette, where the polynucleotide sequence  
CC further comprises control elements capable of driving expression in the  
CC selected host cell, a cell comprising an expression cassette where the  
CC polynucleotide sequence further comprises control elements compatible  
CC with the expression in the cell and a composition for generating an  
CC immunological response, comprising an expression cassette. The expression  
CC cassette and the methods of the invention are useful in eliciting an  
CC immune response, in DNA immunisation, in generation of packaging cell  
CC lines and in producing HIV type C proteins. This sequence represents an  
CC HIV-1 polynucleotide of the invention.

XX SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;  
Query Match 98.3%; Score 2415.4; DB 12; Length 2469;  
Best Local Similarity 99.3%; Pred. No. 1.4e-290;  
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;  
QY 1 GTCCAGCCACCACTGGCGGAGGCCATGAGCAGGCCACACGCGCAACATCTGTATGAG 60  
Db 1 GTCCAGCCACCACTGGCGGAGGCCATGAGCAGGCCACACGCGCAACATCTGTATGAG 60  
QY 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGCG 120  
Db 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGCG 120  
QY 121 CACATCCCGCGCAACTGCGGCGCCCCCGAAGAGGGCTCTGGAAGTGGCGCAAGAG 180  
Db 121 CACATCCCGCGCAACTGCGGCGCCCCCGAAGAGGGCTCTGGAAGTGGCGCAAGAG 180  
QY 181 GGCCACAGATGAAGGACTGCACCGAGCGGCCAGGCCAACTTTCTTCCGCGAGGACCTGGCC 240  
Db 181 GGCCACAGATGAAGGACTGCACCGAGCGGCCAGGCCAACTTTCTTCCGCGAGGACCTGGCC 240  
QY 241 TTCCCGCAGGCAAGGGCCCCGGAGTTCCCGAGCGAGCGAAGCCGCGCCACAGCCCCACC 300  
Db 241 TTCCCGCAGGCAAGGGCCCCGGAGTTCCCGAGCGAGCGAAGCCGCGCCACAGCCCCACC 300  
QY 301 AGCGCGAGCTGCGAGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGAGCGCCAG 360  
Db 301 AGCGCGAGCTGCGAGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGAGCGCCAG 360  
QY 361 GGCACTTGAACCTTCCCGCAGATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG 420  
Db 361 GGCACTTGAACCTTCCCGCAGATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG 420  
QY 421 GGCGGCCAGATCAAGAGGGCCCTCTGTGACACCGCGCGCGAGCACACCTGTGTGAGGAG 480  
Db 421 GGCGGCCAGATCAAGAGGGCCCTCTGTGACACCGCGCGCGAGCACACCTGTGTGAGGAG 480  
QY 481 ATGAGCTGCGCGCAAGTGGAGGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAG 540  
Db 481 ATGAGCTGCGCGCAAGTGGAGGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAG 540  
QY 541 GTGCGCCAGTACACACAGATCTCTGATCGAGATCTGCGGCGAAGAGGCCATCGGCACCGTG 600  
Db 541 GTGCGCCAGTACACACAGATCTCTGATCGAGATCTGCGGCGAAGAGGCCATCGGCACCGTG 600  
QY 601 CTGATCGGCGCCACCGCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCGCCACCGCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAACCTTCCCGCATCGAGCCCGTGCAGACCGTGCCTGGAAGCTGGAAGCCGGCATG 720  
Db 661 ACCCTGAACCTTCCCGCATCGAGCCCGTGCAGACCGTGCCTGGAAGCTGGAAGCCGGCATG 720  
QY 721 GAGCGCCCAAGGTGAAGAGTGGCCCTCTGACCGAGGAGAGATCAAGAGCCCTGACCGCC 780  
Db 721 GAGCGCCCAAGGTGAAGAGTGGCCCTCTGACCGAGGAGAGATCAAGAGCCCTGACCGCC 780  
QY 781 ATCTGCGAGGAGTGAAGAGGAGGCGCAAGATCAAGATCGGCGCCCGAGAACCCCTTAC 840  
Db 781 ATCTGCGAGGAGTGAAGAGGAGGCGCAAGATCAAGATCGGCGCCCGAGAACCCCTTAC 840  
QY 841 AACACCCCGCTGTTCGCCCATCAAGAGAAGAGCAGCACCAAGTGGCGGCAAGCTGGTGAC 900  
Db 841 AACACCCCGCTGTTCGCCCATCAAGAGAAGAGCAGCACCAAGTGGCGGCAAGCTGGTGAC 900  
QY 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGATCCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGATCCCCCAC 960  
QY 961 CCGCGCGGCTGGAAGAGAAGAGAGCGTACCGTGTGAGCGTGGCGGACGCGCTACTTC 1020  
Db 961 CCGCGCGGCTGGAAGAGAAGAGAGCGTACCGTGTGAGCGTGGCGGACGCGCTACTTC 1020



Db 961 CCCGCGGCTGAAGAAGAGCGTGACCGTGTGGAGCTGGCGAGCGCTACTTTC 1020  
 Qy 1021 AGCGTGCCCTGGAGAGACTTTCGGCAAGTACACGGCTTACCATCCCGAGATCAAC 1080  
 Db 1021 AGCGTGCCCTGGAGAGACTTTCGGCAAGTACACGGCTTACCATCCCGAGATCAAC 1080  
 Qy 1081 AACGAGACCCCGGATCCGCTACAGTACAACTGCTGCCCGAGGCTGGAAAGGCGAGC 1140  
 Db 1081 AACGAGACCCCGGATCCGCTACAGTACAACTGCTGCCCGAGGCTGGAAAGGCGAGC 1140  
 Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGGCCCGCAACCCC 1200  
 Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGGCCCGCAACCCC 1200  
 Qy 1201 GAGATCGTATACCA-----GGCCCCCTGTACGTGGGAGCGACCTTGGAGATCGGC 1254  
 Db 1201 GAGATCGTATACCAAGCATGACGACCTGTACGTGGGAGCGACCTTGGAGATCGGC 1260  
 Qy 1255 CAGCACCGGCCAAGTACGAGAGCTGCGCAGCACTGCTGCGCTGGGCTTCAACACC 1314  
 Db 1261 CAGCACCGGCCAAGTACGAGAGCTGCGCAGCACTGCTGCGCTGGGCTTCAACACC 1320  
 Qy 1315 CCCGACAAGACACAGAGAGAGCCCTTCTTCCGCGCAT-----CGAGCTGCAACCCC 1368  
 Db 1321 CCCGACAAGACACAGAGAGAGCCCTTCTTCCGCGCAT-----CGAGCTGCAACCCC 1380  
 Qy 1369 GACAAGTGGACCGTGAGCCCATACGAGTGTCCCGAGAGAGAGCTGGACCGTGAACGAC 1428  
 Db 1381 GACAAGTGGACCGTGAGCCCATACGAGTGTCCCGAGAGAGAGCTGGACCGTGAACGAC 1440  
 Qy 1429 ATCCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCGAGGATCTACCCGGGCATCAAGGTG 1488  
 Db 1441 ATCCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCGAGGATCTACCCGGGCATCAAGGTG 1500  
 Qy 1489 CGCCAGCTGTGAAGCTGTGCGCGCGCCAGAGGCCCTGACCGACATCGTGCCTTGACC 1548  
 Db 1501 CGCCAGCTGTGAAGCTGTGCGCGCGCCAGAGGCCCTGACCGACATCGTGCCTTGACC 1560  
 Qy 1549 GAGGAGGCGAGCTGGAGTGGCCGAGAACCCGCGAGATCTTGGCGAGCCCGTGCAACGGC 1608  
 Db 1561 GAGGAGGCGAGCTGGAGTGGCCGAGAACCCGCGAGATCTTGGCGAGCCCGTGCAACGGC 1620  
 Qy 1609 GTGTACTAGACCCAGCAAGACCTGTGTGGCGGAGATCCAGAACGAGGCCACGACGAG 1668  
 Db 1621 GTGTACTAGACCCAGCAAGACCTGTGTGGCGGAGATCCAGAACGAGGCCACGACGAG 1680  
 Qy 1669 TGGACCTACAGATCTTACAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728  
 Db 1681 TGGACCTACAGATCTTACAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740  
 Qy 1729 ATGCGGACCGCCACACCAACGAGTGAAGCAGTGAACGAGGCCGTGGAGAGATCGCC 1788  
 Db 1741 ATGCGGACCGCCACACCAACGAGTGAAGCAGTGAACGAGGCCGTGGAGAGATCGCC 1800  
 Qy 1789 ATGGAGAGCATCGTGAATCTGGGGCAGACCCCGCAAGTTCGCGCTGCCCATCCAGAGGAG 1848  
 Db 1801 ATGGAGAGCATCGTGAATCTGGGGCAGACCCCGCAAGTTCGCGCTGCCCATCCAGAGGAG 1860  
 Qy 1849 ACCTGGGAGACCTGTGGAGACCTACTTGGCAGGCGACCTGGATCCCGAGTGGGAGTTTC 1908  
 Db 1861 ACCTGGGAGACCTGTGGAGACCTACTTGGCAGGCGACCTGGATCCCGAGTGGGAGTTTC 1920  
 Qy 1909 GTGAACACCCCGCTTGTGAAGCTGTGGTATCCAGTGGAGAAAGGAGCCCATCATTCGGC 1968  
 Db 1921 GTGAACACCCCGCTTGTGAAGCTGTGGTATCCAGTGGAGAGAGAGCCCATCATTCGGC 1980  
 Qy 1969 GCGGAGACCTTCTAGTGGAGCGCGCGCCACCGGAGACCAAGATCGGCAAGGCCGCGC 2028  
 Db 1981 GCGGAGACCTTCTAGTGGAGCGCGCGCGCCACCGGAGACCAAGATCGGCAAGGCCGCGC 2040  
 Qy 2029 TACGTGACCGACCGGGCGCGGAGAGATCGTGAGCTTGACCGAGACCAACCAACGAGAG 2088  
 Db 2041 TACGTGACCGACCGGGCGCGGAGAGATCGTGAGCTTGACCGAGACCAACCAACGAGAG 2100

Qy 2089 ACCGAGCTGCAGGCCATCCAGCTGSCCTGCAGGACAGCGGAGGTGAACATCGTG 2148  
 Db 2101 ACCGAGCTGCAGGCCATCCAGCTGSCCTGCAGGACAGCGGAGGTGAACATCGTG 2160  
 Qy 2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCGACAGAGGAGAGCGAG 2208  
 Db 2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCGACAGAGGAGAGCGAG 2220  
 Qy 2209 CTGTGAAACCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACTGAGCTGGGTG 2268  
 Db 2221 CTGTGAAACCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACTGAGCTGGGTG 2280  
 Qy 2269 CCGCCCCACAAGGGCATCGCGGCAACGAGCAGATCGACAGCTGGTGCAGCAAGGGCATC 2328  
 Db 2281 CCGCCCCACAAGGGCATCGCGGCAACGAGCAGATCGACAGCTGGTGCAGCAAGGGCATC 2340  
 Qy 2329 CGCAAGGTGCTGTTCCTGGAGCGCATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2388  
 Db 2341 CGCAAGGTGCTGTTCCTGGAGCGCATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2400  
 Qy 2389 GACCTGTACGTGGGAGCGCGGCCCTTAGGATCGAATTAAGCTTCCCGGGCTTAGCACC 2448  
 Db 2401 GACCTGTACGTGGGAGCGCGGCCCTTAGGATCGAATTAAGCTTCCCGGGCTTAGCACC 2460  
 Qy 2449 GGTGAATTC 2457  
 Db 2461 GGTGAATTC 2469

Search completed: June 1, 2005, 11:33:34  
 Job time : 1272.25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 10:29:57 ; Search time 1407.9 Seconds  
(without alignments)  
10728.225 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457

Sequence: 1 gtcagccaccatggccga.....gggtagcaccggtgaattc 2457

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5706582 seqs, 3073711274 residues

Total number of hits satisfying chosen parameters: 11413164

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2457	100.0	2457	10	US-09-899-575-32
2	2443.4	99.4	2445	15	US-10-190-435-43
3	2443.4	99.4	2445	16	US-10-190-305A-37
4	2436.2	99.2	2463	10	US-10-190-435-9
5	2436.2	99.2	2463	15	US-09-899-575-31
6	2434.8	99.1	2430	10	US-10-190-435-10
7	2434.8	99.1	2430	15	US-10-190-435-11
8	2434.8	99.1	5184	15	US-10-190-435-58
9	2434.8	99.1	5184	16	US-10-190-305A-82
10	2428.6	98.8	2457	15	US-10-190-435-44
11	2428.6	98.8	2457	16	US-10-190-305A-38

12	2415.4	98.3	2469	10	US-09-899-575-30
13	2404.4	97.9	3531	15	US-10-190-435-13
14	2403.4	97.8	2457	15	US-10-190-435-45
15	2403.4	97.8	2457	16	US-10-190-305A-39
16	2402.8	97.8	3537	15	US-10-190-435-14
17	2402.8	97.8	3537	15	US-10-190-435-15
18	2401.8	97.8	5145	15	US-10-190-435-12
19	2401.8	97.8	5145	16	US-10-190-305A-12
20	2343.4	95.4	3607	15	US-10-190-435-48
21	2343.4	95.4	3607	16	US-10-190-305A-42
22	2325.2	94.6	3597	15	US-10-190-435-46
23	2325.2	94.6	3597	16	US-10-190-305A-40
24	2295.8	93.4	3624	15	US-10-190-435-47
25	2295.8	93.4	3624	16	US-10-190-305A-41
26	2136	86.9	2460	16	US-10-241-009-30
27	2136	86.9	2460	16	US-10-190-434B-30
28	2136	86.9	2460	16	US-10-190-305A-32
29	2135	86.9	3564	16	US-10-241-009-13
30	2135	86.9	3564	16	US-10-190-434B-13
31	2135	86.9	3564	16	US-10-190-434B-14
32	2135	86.9	3564	16	US-10-190-435-17
33	2134.4	86.9	4716	15	US-10-190-435-17
34	2134.4	86.9	4716	16	US-10-190-305A-13
35	2130.6	86.7	3999	16	US-10-241-009-9
36	2130.6	86.7	3999	16	US-10-190-434B-9
37	2129	86.7	3999	16	US-10-241-009-10
38	2129	86.7	3999	16	US-10-241-009-11
39	2129	86.7	3999	16	US-10-190-434B-10
40	2129	86.7	3999	16	US-10-190-434B-11
41	2129	86.7	5283	16	US-10-241-009-54
42	2129	86.7	5283	16	US-10-190-434B-54
43	2129	86.7	5283	16	US-10-190-305A-75
44	2128.8	86.6	4713	15	US-10-190-435-59
45	2128.8	86.6	4713	16	US-10-190-305A-83

#### ALIGNMENTS

#### RESULT 1

US-09-899-575-32  
; Sequence 32, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megede, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Estrelita Janse  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; CURRENT FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 09/475,704  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 32  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Description of Artificial Sequence: PR975YMMW  
US-09-899-575-32

Query Match 100.0%; Score 2457; DB 10; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GTGAGCCACCATGGCCGAGCCATGAGCCAGCCAGCCAGCCACATCTCTGATGAG 60  
Db 1 GTGAGCCACCATGGCCGAGCCATGAGCCAGCCAGCCAGCCACATCTCTGATGAG 60

QY 61 CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAA CTGGGCAAGAGGCG 120  
Db 61 CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAA CTGGGCAAGAGGCG 120  
QY 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGGCTGCTGGAAGTGGGCAAGAG 180  
Db 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGGCTGCTGGAAGTGGGCAAGAG 180  
QY 181 GGCACACGATCAAGGATGTGACACGAGCGCCAGGCCAACTTCTTCGGGAGCACTGGCC 240  
Db 181 GGCACACGATCAAGGATGTGACACGAGCGCCAGGCCAACTTCTTCGGGAGCACTGGCC 240  
QY 241 TTCCCCCAAGGGCAAGGCCCGCGAGTTCCACGAGCAGAAACCGCGCCCAACAGCCCCAC 300  
Db 241 TTCCCCCAAGGGCAAGGCCCGCGAGTTCCACGAGCAGAAACCGCGCCCAACAGCCCCAC 300  
QY 301 AGCCGCGAGCTGACAGTGGCGGCGCAAAACCCCGCAGCGAGCGCGCGCCGAGCGCCAG 360  
Db 301 AGCCGCGAGCTGACAGTGGCGGCGCAAAACCCCGCAGCGAGCGCGCGCCGAGCGCCAG 360  
QY 361 GGCACCTGAACTTCCCCAGATCACCTGTGGGAGCGCCCCCTGGTGAGCATCAAGTG 420  
Db 361 GGCACCTGAACTTCCCCAGATCACCTGTGGGAGCGCCCCCTGGTGAGCATCAAGTG 420  
QY 421 GGCAGCCAGATCAAGAGGCCCTGCTGGACACCGCGCGCAGCACACGTCGTGAGGAG 480  
Db 421 GGCAGCCAGATCAAGAGGCCCTGCTGGACACCGCGCGCAGCACACGTCGTGAGGAG 480  
QY 481 ATGAGCTTCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAACAAG 540  
Db 481 ATGAGCTTCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAACAAG 540  
QY 541 GTGGCCAGTAGCAGACAGATCCTGATCGAGATCTGGGCAAGAGGCCATCGGCAACG 600  
Db 541 GTGGCCAGTAGCAGACAGATCCTGATCGAGATCTGGGCAAGAGGCCATCGGCAACG 600  
QY 601 CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720  
Db 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720  
QY 721 GAGGCCCCAGGTGACGAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
Db 721 GAGGCCCCAGGTGACGAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
QY 781 ATCTCGGAGGAGATGGAAGAGGCGCAAGATCAACAAGATCGGCCCCGAGAAACCCCTAC 840  
Db 781 ATCTCGGAGGAGATGGAAGAGGCGCAAGATCAACAAGATCGGCCCCGAGAAACCCCTAC 840  
QY 841 AACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCAACAAGTGGCGCAAGCTGGTGAC 900  
Db 841 AACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCAACAAGTGGCGCAAGCTGGTGAC 900  
QY 901 TTCGCGAGCTGAACAGGCAACCCAGACTTCTGGGAGTGCAGCTGGGCACTCCCCAC 960  
Db 901 TTCGCGAGCTGAACAGGCAACCCAGACTTCTGGGAGTGCAGCTGGGCACTCCCCAC 960  
QY 961 CCCGCGGCTCAAGAAAGAGAGCGTGACCGTGTGACGCTGGGCGAGCCCTACTTC 1020  
Db 961 CCCGCGGCTCAAGAAAGAGAGCGTGACCGTGTGACGCTGGGCGAGCCCTACTTC 1020  
QY 1021 AGCGTCCCCCTGGAGGAGACTTCGCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
Db 1021 AGCGTCCCCCTGGAGGAGACTTCGCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
QY 1081 AACGAGACCCCGGATCCGCTACCAAGTACAAAGTGTGCTGCCCGAGGCTGGAGGCGAGC 1140  
Db 1081 AACGAGACCCCGGATCCGCTACCAAGTACAAAGTGTGCTGCCCGAGGCTGGAGGCGAGC 1140  
QY 1141 CCCAGCATCTTCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCGCGCGCCGCAACCCC 1200

Db 1141 CCCAGCATCTTCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCGCGCGCCGCAACCCC 1200  
QY 1201 GAGATCGTGATCTACAGGCCCCCTCTGTACTGTGGGACGAGCCTTGGAGATCGGCCAGCAC 1260  
Db 1201 GAGATCGTGATCTACAGGCCCCCTCTGTACTGTGGGACGAGCCTTGGAGATCGGCCAGCAC 1260  
QY 1261 CGGCCAAGATCGAGGAGCTGCGAAGCACCTGTGCGCTGGGGCTTCAACACCCCGGAC 1320  
Db 1261 CGGCCAAGATCGAGGAGCTGCGAAGCACCTGTGCGCTGGGGCTTCAACACCCCGGAC 1320  
QY 1321 AAGAAGCACCAAGAGGAGCCCTTCTGCTGCCATCGAGCTGCAACCCGCAAGTGGACC 1380  
Db 1321 AAGAAGCACCAAGAGGAGCCCTTCTGCTGCCATCGAGCTGCAACCCGCAAGTGGACC 1380  
QY 1381 GTGACGCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440  
Db 1381 GTGACGCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440  
QY 1441 GTGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGTGGCCGCTGTGC 1500  
Db 1441 GTGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGTGGCCGCTGTGC 1500  
QY 1501 AAGCTGCTGCGCGCGCCCAAGGCCCTGACCGACATCTGTGCCCTGACCGAGGAGCCGAG 1560  
Db 1501 AAGCTGCTGCGCGCGCCCAAGGCCCTGACCGACATCTGTGCCCTGACCGAGGAGCCGAG 1560  
QY 1561 CTGAGCTGGCGCGAGAACCGGAGATCTGTGCGGAGCCCGTGCAACGCGTGTACTACGAC 1620  
Db 1561 CTGAGCTGGCGCGAGAACCGGAGATCTGTGCGGAGCCCGTGCAACGCGTGTACTACGAC 1620  
QY 1621 CCCAGAAAGACCTGCTGGCGCAGATCCAGAGCAGGCGCCAGCAGCTGAGCTTACACAG 1680  
Db 1621 CCCAGAAAGACCTGCTGGCGCAGATCCAGAGCAGGCGCCAGCAGCTGAGCTTACACAG 1680  
QY 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGTGCACCGCC 1740  
Db 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGTGCACCGCC 1740  
QY 1741 CACACAAAGAGTGAAGAGCTGACCGAGCGCTGAGAGAGATCGCCATGGAGAGCATC 1800  
Db 1741 CACACAAAGAGTGAAGAGCTGACCGAGCGCTGAGAGAGATCGCCATGGAGAGCATC 1800  
QY 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTTGGAGAGCC 1860  
Db 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTTGGAGAGCC 1860  
QY 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCC 1920  
Db 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCC 1920  
QY 1921 CCCCTGGTGAAGCTGTGGTACAGCTGGAGAAAGAGCCCATCATCGGCGCCGAGACCTTC 1980  
Db 1921 CCCCTGGTGAAGCTGTGGTACAGCTGGAGAAAGAGCCCATCATCGGCGCCGAGACCTTC 1980  
QY 1981 TACGTGGACGCGCGCCCAACCGCAGACCAAGATTCGGCAAGCCGCTTACGTGACCCAG 2040  
Db 1981 TACGTGGACGCGCGCCCAACCGCAGACCAAGATTCGGCAAGCCGCTTACGTGACCCAG 2040  
QY 2041 CGGGCCCGCGCAGAAAGATCGTGAGCCTGACCGAGACCAACAACAGAGAGCCGAGCTGAC 2100  
Db 2041 CGGGCCCGCGCAGAAAGATCGTGAGCCTGACCGAGACCAACAACAGAGAGCCGAGCTGAC 2100  
QY 2101 GGCATCAGCTGGCCCTGAGAGACGCGGACGAGGAGTGAACATCGTGACCGAGCCAG 2160  
Db 2101 GGCATCAGCTGGCCCTGAGAGACGCGGACGAGGAGTGAACATCGTGACCGAGCCAG 2160  
QY 2161 TACGCCCTGGGATCATCCAGGCCAGCCCGCAAGAGCGAGCGAGCTGTGTAACCCAG 2220  
Db 2161 TACGCCCTGGGATCATCCAGGCCAGCCCGCAAGAGCGAGCGAGCTGTGTAACCCAG 2220  
QY 2221 ATCTCAGGACGCTGATCAAGAGGAGAGGTTGTAACCTGAGCTGGGTGC CGGCCCAACAG 2280



```

Db 1501 CTGCGCGGCCCAAGCCCTGACGACATCGTGCCCTGACCGAGGAGCGAGCTGGAG 1560
QY 1567 CTGCGCGAGAACCGGAGATCTGCGCGAGCCGCTGACCGGCTGTACTACAGCCCGAGC 1626
Db 1561 CTGCGCGAGAACCGGAGATCTGCGCGAGCCGCTGACCGGCTGTACTACAGCCCGAGC 1620
QY 1627 AAGGACCTGGTGGCGGAGATCCAGAAAGCAGAGGCGCCACGACCTGACCTTACCAGATCTAC 1686
Db 1621 AAGGACCTGGTGGCGGAGATCCAGAAAGCAGAGGCGCCACGACCTGACCTTACCAGATCTAC 1680
QY 1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCGACCGCCACACC 1746
Db 1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCGACCGCCACACC 1740
QY 1747 AACGAGCTGAAGCAGCTGACCGAGCGCTGCGAGAGATCGCCATCGAGAGCATCTGTATC 1806
Db 1741 AACGAGCTGAAGCAGCTGACCGAGCGCTGCGAGAGATCGCCATCGAGAGCATCTGTATC 1800
QY 1807 TGGGCGAAGACCCCAAGTTCCGCTGCGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGG 1866
Db 1801 TGGGCGAAGACCCCAAGTTCCGCTGCGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGG 1860
QY 1867 ACCGACTACTGCGAGGCCACCTGGATCCCGGAGTGGGAGTTTGGTGAACACACCCCGCCCTG 1926
Db 1861 ACCGACTACTGCGAGGCCACCTGGATCCCGGAGTGGGAGTTTGGTGAACACACCCCGCCCTG 1920
QY 1927 GTGAAGCTGTGTATCAGCTGTGAGAAAGAGGCCCATCATCGGCGCGAGACCTTCTACGTG 1986
Db 1921 GTGAAGCTGTGTATCAGCTGTGAGAAAGAGGCCCATCATCGGCGCGAGACCTTCTACGTG 1980
QY 1987 GAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGGC 2046
Db 1981 GAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGGC 2040
QY 2047 CGCGAGAAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGCGCATC 2106
Db 2041 CGCGAGAAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGCGCATC 2100
QY 2107 GAGCTGGCCCTCAGGACAGCGGCGAGCGAGTGAAATCATGTCACCGACAGCCAGTACGCC 2166
Db 2101 GAGCTGGCCCTCAGGACAGCGGCGAGCGAGTGAAATCATGTCACCGACAGCCAGTACGCC 2160
QY 2167 CTGGGCATCATCAGCGCCAGCCGACAAAGAGCGAGCGAGCTGGTGAACCAAGATCATC 2226
Db 2161 CTGGGCATCATCAGCGCCAGCCGACAAAGAGCGAGCGAGCTGGTGAACCAAGATCATC 2220
QY 2227 GAGCAGCTGATCAAGAAGAGAGGTGTACTGAGCTGGGTGCCCGCCCAACAAAGGCGATC 2286
Db 2221 GAGCAGCTGATCAAGAAGAGAGGTGTACTGAGCTGGGTGCCCGCCCAACAAAGGCGATC 2280
QY 2287 GCGCGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGGATCGCAAGGTGCTGTCTGT 2346
Db 2281 GCGCGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGGATCGCAAGGTGCTGTCTGT 2340
QY 2347 GAGCGCATCGATGGCGGATCGTATCTACAGTACATGAGCAGCTGTACGTGGGCGAGC 2406
Db 2341 GAGCGCATCGATGGCGGATCGTATCTACAGTACATGAGCAGCTGTACGTGGGCGAGC 2400
QY 2407 GCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db 2401 GCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2445

```

RESULT 3

```

US-10-190-305A-37
; Sequence 37, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: EUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

```

```

; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190.305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: p2Pol.opt.YMMW_C
US-10-190-305A-37

Query Match 99.4%; Score 2443.4; DB 16; Length 2445;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 GCCACCATGGCGGAGGCCATGAGCCAGCCAGCCAGCCATCTCAAGTGTCTTCAACTGCGCAAGAGGGCCACATC 66
Db 1 GCCACCATGGCGGAGGCCATGAGCCAGCCAGCCAGCCAGCCATCTCAAGTGTCTTCAACTGCGCAAGAGGGCCAC 60
QY 67 AACTTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGCCACATC 126
Db 61 AACTTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGCCACATC 120
QY 127 GCCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCAC 186
Db 121 GCCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCAC 180
QY 187 CAGATGAAGGACTGCAACCGAGCGCGCAGCCAACTTCTTCGCGAGGACTGTGGCTTTCCCG 246
Db 181 CAGATGAAGGACTGCAACCGAGCGCGCAGCCAACTTCTTCGCGAGGACTGTGGCTTTCCCG 240
QY 247 CAGGCAAGGGCCCGCGAGTTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCGCCACAGCGCGC 306
Db 241 CAGGCAAGGGCCCGCGAGTTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCGCCACAGCGCGC 300
QY 307 GAGCTGAGGTGGCGGCGGACAAACCCCGAGCGAGGCGCGGCGCGAGCGCGCGAGGGCCAGGCGC 366
Db 301 GAGCTGAGGTGGCGGCGGACAAACCCCGAGCGAGGCGCGGCGCGAGCGCGCGAGGGCCAGGCGC 360
QY 367 CTGAATCTTCCCGCAGATCAGCTGTGGCAGCGCCCTTGGTGAGCATCAAGGTGGCGCGC 426
Db 361 CTGAATCTTCCCGCAGATCAGCTGTGGCAGCGCCCTTGGTGAGCATCAAGGTGGCGCGC 420
QY 427 CAGATCAAGGAGGGCCCTCTGAGCACCGCGCGCAGACACCGTGTCTGGAGGAGATGAGC 486
Db 421 CAGATCAAGGAGGGCCCTCTGAGCACCGCGCGCAGACACCGTGTCTGGAGGAGATGAGC 480
QY 487 CTGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGGCGGCTTTCATCAAGGTGGCG 546
Db 481 CTGCGCGCAAGTGGAGCCCAAGATGATGATCGCGCGCATCGGCGGCTTTCATCAAGGTGGCG 540
QY 547 CAGTACGACCAAGATCTCTGATCGAGATCTGCGCAAGAGGCGCATCGGCAACCGTGTCTGATC 606
Db 541 CAGTACGACCAAGATCTCTGATCGAGATCTGCGCAAGAGGCGCATCGGCAACCGTGTCTGATC 600
QY 607 GCGCCCAACCCCGTGAAACATCATCTGCGCGCAACATGCTGACCCAGCTGGGTGCAACCTG 666
Db 601 GCGCCCAACCCCGTGAAACATCATCTGCGCGCAACATGCTGACCCAGCTGGGTGCAACCTG 660
QY 667 AACTTCCCGCATAGCCCGCATCGAGCCCGTGGTGAAGCTTGAAGCCCGGCGCATGAGCGCG 726
Db 661 AACTTCCCGCATAGCCCGCATCGAGCCCGTGGTGAAGCTTGAAGCCCGGCGCATGAGCGCG 720
QY 727 CCCAAGGTGAAGAGTGGCCCTCTGACCGAGGAGAGATCAAGGCGCTTGCACCGCCATCTGC 786
Db 721 CCCAAGGTGAAGAGTGGCCCTCTGACCGAGGAGAGATCAAGGCGCTTGCACCGCCATCTGC 780
QY 787 GAGGAGATGGAGAGGAGGCGAAGATCACCAAGATCGCGCCCGCGAGAAACCCCTTACAACACC 846
Db 781 GAGGAGATGGAGAGGAGGCGAAGATCACCAAGATCGCGCCCGCGAGAAACCCCTTACAACACC 840

```





QY	134	ACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAGGGGCCACCAAGATGA	193
Db	1607	ACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAGGGGCCACCAAGATGA	1666
QY	194	AGGACTGCACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACCTGGCTTCCCGCCAGGGCA	253
Db	1667	AGGACTGCACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACCTGGCTTCCCGCCAGGGCA	1726
QY	254	AGGCGCGAGTTCCCGAGCGAGCGAACCAGCGCCCAACAGCCCCCAACAGCGCGAGCTGC	313
Db	1727	AGGCGCGAGTTCCCGAGCGAGCGAACCAGCGCCCAACAGCCCCCAACAGCGCGAGCTGC	1786
QY	314	AGGTGCGCGGACCAACCCCCCGAGGCGCGCGCGAGCGCGAGGCGCACCTGGAAT	373
Db	1787	AGGTGCGCGGACCAACCCCCCGAGGCGCGCGCGAGGCGCGAGGCGCACCTGGAAT	1846
QY	374	TCCCCAGATCACTTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGCGCGCAGATCA	433
Db	1847	TCCCCAGATCACTTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGCGCGCAGATCA	1906
QY	434	AGGAGCCCTGTGACACCGCGCGCGAGCAACCGTGTGAGAGAGATGAGCCTGCCG	493
Db	1907	AGGAGCCCTGTGAGACACCGCGCGCGAGCAACCGTGTGAGAGAGATGAGCCTGCCG	1966
QY	494	GCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGGCTTCATCAAGGTGGCCAGTAGG	553
Db	1967	GCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGGCTTCATCAAGGTGGCCAGTAGG	2026
QY	554	ACCAGATCCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA	613
Db	2027	ACCAGATCCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA	2086
QY	614	CCCCGTGAACATCATTCGGCCGCAATGCTGACCCAGCTGGGTGCAACCTTGAATTC	673
Db	2087	CCCCGTGAACATCATTCGGCCGCAATGCTGACCCAGCTGGGTGCAACCTTGAATTC	2146
QY	674	CCATCAGCCCCCATCGAGACCGTGGCCGTAAGCTGNAAGCCGGCATGACCGCCCAAGG	733
Db	2147	CCATCAGCCCCCATCGAGACCGTGGCCGTAAGCTGNAAGCCGGCATGACCGCCCAAGG	2206
QY	734	TGAAGCAGTGGCCCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCCATCTCGAGGAGA	793
Db	2207	TGAAGCAGTGGCCCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCCATCTCGAGGAGA	2266
QY	794	TGGAGAAGAGGGCAAGATFACAAAGATGCGCCCCGAGAAACCCCTACAAACACCCCGTGT	853
Db	2267	TGGAGAAGAGGGCAAGATFACAAAGATGCGCCCCGAGAAACCCCTACAAACACCCCGTGT	2326
QY	854	TGCGCATCAAGAAGAGGACAGCAACAGTGGCGCAAGCTGGTGACTTCGCGAGCTGA	913
Db	2327	TGCGCATCAAGAAGAGGACAGCAACAGTGGCGCAAGCTGGTGACTTCGCGAGCTGA	2386
QY	914	ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAACCCCGCGCCTGA	973
Db	2387	ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAACCCCGCGCCTGA	2446
QY	974	AGAAGAAGAGCGTGCCTGTGGAGCTGGGCGACGCCCTACTTCAGCGTGGCCCTGG	1033
Db	2447	AGAAGAAGAGCGTGCCTGTGGAGCTGGGCGACGCCCTACTTCAGCGTGGCCCTGG	2506
QY	1034	ACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACCCCG	1093
Db	2507	ACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACCCCG	2566
QY	1094	GCATCCGCTACCACTACAGCTGCTGCCCGAGGGCTGGAAGGGCAGCCCCAGCATCTTCC	1153
Db	2567	GCATCCGCTACCACTACAGCTGCTGCCCGAGGGCTGGAAGGGCAGCCCCAGCATCTTCC	2626
QY	1154	AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGCGCGCAACCCCGAGATCTGTATCT	1213
Db	2627	AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGCGCGCAACCCCGAGATCTGTATCT	2686

QY	1214	ACCAGGCCCCCTGTACTGTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCAAGATCG	1273
Db	2687	ACCAGGCCCCCTGTACTGTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCAAGATCG	2746
QY	1274	AGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTACCAACCCCGCAAGAAGACCAAGA	1333
Db	2747	AGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTACCAACCCCGCAAGAAGACCAAGA	2806
QY	1334	AGGAGCCCCCTTCTGCTGCCATCGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCG	1393
Db	2807	AGGAGCCCCCTTCTGCTGCCATCGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCG	2866
QY	1394	AGCTGCCCGAAGAGGAGAGCTGGAACCGTGAACGATCCAGAAGCTGTGGGCAAGCTGA	1453
Db	2867	AGCTGCCCGAAGAGGAGAGCTGGAACCGTGAACGATCCAGAAGCTGTGGGCAAGCTGA	2926
QY	1454	ACTGGGCGAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGG	1513
Db	2927	ACTGGGCGAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGG	2986
QY	1514	GCGCCAAAGCCCTGACCGACATCGTGCCCCCTGACCGAGAGCGCGAGCTGGAGCTGGCG	1573
Db	2987	GCGCCAAAGCCCTGACCGACATCGTGCCCCCTGACCGAGAGCGCGAGCTGGAGCTGGCG	3046
QY	1574	AGAACCGGAGATCTCTGCGCGAGCCGTCACGGCGTGTACTACGACCCCGCAGCAAGGACC	1633
Db	3047	AGAACCGGAGATCTCTGCGCGAGCCGTCACGGCGTGTACTACGACCCCGCAGCAAGGACC	3106
QY	1634	TGTTGGCGGAGATCCAGAAGCAGGCGCCACGACCAAGTGGACCTTACAGATCTTACAGGAGC	1693
Db	3107	TGTTGGCGGAGATCCAGAAGCAGGCGCCACGACCAAGTGGACCTTACAGATCTTACAGGAGC	3166
QY	1694	CCTTCAAGAACCTGAAAGACCGGCAAGTAGGCAAGATCGGCATCGAGAGCATCTGTATCTGGGCA	1753
Db	3167	CCTTCAAGAACCTGAAAGACCGGCAAGTAGGCAAGATCGGCATCGAGAGCATCTGTATCTGGGCA	3226
QY	1754	TGAAGCAGCTCACCGAGGCGCTGCAAGAAGATCGCCATCGAGAGCATCTGTATCTGGGCA	1813
Db	3227	TGAAGCAGCTCACCGAGGCGCTGCAAGAAGATCGCCATCGAGAGCATCTGTATCTGGGCA	3286
QY	1814	AGACCCCAAGTTTCGCTGCTCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT	1873
Db	3287	AGACCCCAAGTTTCGCTGCTCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT	3346
QY	1874	ACTGGCAGGCACCTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGC	1933
Db	3347	ACTGGCAGGCACCTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGC	3406
QY	1934	TGTGTACAGCTGGAGAAGGAGGCCATCATCGCGCGCGAGACCTTCTACCTGGACGGCG	1993
Db	3407	TGTGTACAGCTGGAGAAGGAGGCCATCATCGCGCGCGAGACCTTCTACCTGGACGGCG	3466
QY	1994	CCGCCAACCCGGGAGACCAAGATCGGCAAGGCGGCTACGTGAACCGACCGGGCGGGCAGA	2053
Db	3467	CCGCCAACCCGGGAGACCAAGATCGGCAAGGCGGCTACGTGAACCGACCGGGCGGGCAGA	3526
QY	2054	AGATCGTGAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCGCATTCAGCTGG	2113
Db	3527	AGATCGTGAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCGCATTCAGCTGG	3586
QY	2114	CCCTGCAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCGACAGCGCAGTACCGCTGGGCA	2173
Db	3587	CCCTGCAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCGACAGCGCAGTACCGCTGGGCA	3646
QY	2174	TCATCCAGGCCAGCCCGACAAAGCAGCGAGCGAGCTGGTGAACACAGATCATTCGAGCAGC	2233
Db	3647	TCATCCAGGCCAGCCCGACAAAGCAGCGAGCGAGCTGGTGAACACAGATCATTCGAGCAGC	3706
QY	2234	TGATCAAGAAGAGAGGTGTACTCTGAGCTGGGTGCCCGCCCAACAGGGGCATTCGCGGCA	2293
Db	3707	TGATCAAGAAGAGAGGTGTACTCTGAGCTGGGTGCCCGCCCAACAGGGGCATTCGCGGCA	3766
QY	2294	ACGAGCAGATCGAACAAAGCTGTGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTCTGGAGCGCA	2353

Db 3767 ACAGCAGATCGACAAAGCTGTGGAGCAAGGGCATCCGAAGGTGCTGTTCCTGGACGGCA 3826  
 Qy 2354 TCGATGGCGGATCGTGTATCTACCAAGTACATGGAGACCTGTACGTGGGAGCGCGGCC 2413  
 Db 3827 TCGATGGCGGATCGTGTATCTACCAAGTACATGGAGACCTGTACGTGGGAGCGCGGCC 3886  
 Qy 2414 CTAGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2451  
 Db 3887 CTAGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 3924

RESULT 5

US-09-899-575-31  
 ; Sequence 31, Application US/09899575  
 ; Publication No. US2003023961A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zur Megede, Jan  
 ; APPLICANT: Barnett, Susan W.  
 ; APPLICANT: Egnelbrecht, Susan  
 ; APPLICANT: van Rensburg, Estrelita Janse  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
 ; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
 ; FILE REFERENCE: PP01631.102  
 ; CURRENT APPLICATION NUMBER: US/09/899,575  
 ; CURRENT FILING DATE: 2001-07-05  
 ; PRIOR APPLICATION NUMBER: 09/475,704  
 ; PRIOR FILING DATE: 1999-12-30  
 ; NUMBER OF SEQ ID NOS: 135  
 ; SOFTWARE: Patent in Ver. 2.0  
 ; SEQ ID NO 31  
 ; LENGTH: 2463  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: PR975YM  
 US-09-899-575-31

Query Match 99.2%; Score 2436.2; DB 10; Length 2463;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
 Qy 1 GTCAGCCACCATGGCGAGGCCATGAGCGAGCCACCGAGCGCAACATCTCTGATGAG 60  
 Db 1 GTCAGCCACCATGGCGAGGCCATGAGCGAGCCACCGAGCGCAACATCTCTGATGAG 60  
 Qy 61 CGCAGCAACTTCAAGGGCCCCAAGGGCATCATCAAGTCTTCAACTGGGCAAGAGGGC 120  
 Db 61 CGCAGCAACTTCAAGGGCCCCAAGGGCATCATCAAGTCTTCAACTGGGCAAGAGGGC 120  
 Qy 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
 Db 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
 Qy 181 GGCCACCAAGTGAAGGACTGACCGAGCGCCAGCGCCAACTTCTTCGCGAGGACTTGCC 240  
 Db 181 GGCCACCAAGTGAAGGACTGACCGAGCGCCAGCGCCAACTTCTTCGCGAGGACTTGCC 240  
 Qy 241 TTCCCCCAAGGAGGCGCGGAGTTTCCAGAGCGAGCAAAACCGCGGCAACAGCCCCACC 300  
 Db 241 TTCCCCCAAGGAGGCGCGGAGTTTCCAGAGCGAGCAAAACCGCGGCAACAGCCCCACC 300  
 Qy 301 AGCCGCGAGCTGAGGTTGCGGCGCAACACCCCGAGCGAGGCGCGGCGCGAGCGCCAG 360  
 Db 301 AGCCGCGAGCTGAGGTTGCGGCGCAACACCCCGAGCGAGGCGCGGCGCGAGCGCCAG 360  
 Qy 361 GGCAACCTTGAACCTTCCCGGAGATCACTGTGGGAGCGCCCTCTGGTGAAGCATCAAGTG 420  
 Db 361 GGCAACCTTGAACCTTCCCGGAGATCACTGTGGGAGCGCCCTCTGGTGAAGCATCAAGTG 420  
 Qy 421 GGCGGCGAGATCAAGGAGGCGCTGCTGGACACCGCGCGCGACACCGTCTGGAGGAG 480  
 Db 421 GGCGGCGAGATCAAGGAGGCGCTGCTGGACACCGCGCGCGACACCGTCTGGAGGAG 480

Qy 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG 540  
 Db 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG 540  
 Qy 541 GTGCGCGATGACGACGAGATCTGATGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
 Db 541 GTGCGCGATGACGACGAGATCTGATGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
 Qy 601 CTGATCGGCGCCACCCCGCGTGAACATCATCGGCGCGCAACATGCTGAGCCAGCTGGGCTGC 660  
 Db 601 CTGATCGGCGCCACCCCGCGTGAACATCATCGGCGCGCAACATGCTGAGCCAGCTGGGCTGC 660  
 Qy 661 ACCCTGAACCTTCCCGCATCAGCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGCGATG 720  
 Db 661 ACCCTGAACCTTCCCGCATCAGCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGCGATG 720  
 Qy 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTTGACGAGAGAGATCAAGGCGCTTGAACCGCC 780  
 Db 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTTGACGAGAGAGATCAAGGCGCTTGAACCGCC 780  
 Qy 781 ATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCGCGCGAAGCCCTTAC 840  
 Db 781 ATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCGCGCGAAGCCCTTAC 840  
 Qy 841 AACACCCCGCTGTTCGCCATCAAGAGAGAGAGACCAAGTGGCGCGCAAGCTGGTGGAC 900  
 Db 841 AACACCCCGCTGTTCGCCATCAAGAGAGAGAGACCAAGTGGCGCGCAAGCTGGTGGAC 900  
 Qy 901 TTCCGCGAGTGAACAAGCGCACCCAGGACTTTCGGAGGTGACGCTGGGCGATCCCCCAC 960  
 Db 901 TTCCGCGAGTGAACAAGCGCACCCAGGACTTTCGGAGGTGACGCTGGGCGATCCCCCAC 960  
 Qy 961 CCGCGCGCTGAGAGAGAGAGAGCGTGAACCGCTGCGAGCTGGGCGAGCGCTTACTTTC 1020  
 Db 961 CCGCGCGCTGAGAGAGAGAGAGCGTGAACCGCTGCGAGCTGGGCGAGCGCTTACTTTC 1020  
 Qy 1021 AGCGTGGCCCTTGACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCCCAGCATCAAC 1080  
 Db 1021 AGCGTGGCCCTTGACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCCCAGCATCAAC 1080  
 Qy 1081 AACGAGACCCCGGCGATCCGCTACCAAGTACAAAGTGTGCGGCGAGGCTGGAAGGGCAGC 1140  
 Db 1081 AACGAGACCCCGGCGATCCGCTACCAAGTACAAAGTGTGCGGCGAGGCTGGAAGGGCAGC 1140  
 Qy 1141 CCGAGCATCTTCCAGAGCAGATGACCAAGATCTGAGGCGCTTCCGCGCGCGCAACCCC 1200  
 Db 1141 CCGAGCATCTTCCAGAGCAGATGACCAAGATCTGAGGCGCTTCCGCGCGCGCAACCCC 1200  
 Qy 1201 GAGATCGTGTATCTACAGGCGCCCTCTGTAGTGGGCGAGCGACCTGGAGATCGGCGAGCAG 1260  
 Db 1201 GAGATCGTGTATCTACAGGCGCCCTCTGTAGTGGGCGAGCGACCTGGAGATCGGCGAGCAG 1260  
 Qy 1261 CGCGCCAAAGTTCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGCTTTCACACCCCCGAG 1320  
 Db 1261 CGCGCCAAAGTTCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGCTTTCACACCCCCGAG 1320  
 Qy 1321 AAGAGCACAGAGAGGAGGCGCCCTTCTGCGCAT-----CGAGCTGCAACCCCGAG 1374  
 Db 1321 AAGAGCACAGAGAGGAGGCGCCCTTCTGCGCAT-----CGAGCTGCAACCCCGAG 1374  
 Qy 1375 TGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGAGTGGACCGGTGAACGATCCAG 1434  
 Db 1375 TGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGAGTGGACCGGTGAACGATCCAG 1434  
 Qy 1435 AAGCTGTGGGCGCAAGCTGAAGCTGGGCGAGCGAGATCTACCCCGGCGATCAAGTGGCGCAG 1494  
 Db 1435 AAGCTGTGGGCGCAAGCTGAAGCTGGGCGAGCGAGATCTACCCCGGCGATCAAGTGGCGCAG 1494  
 Qy 1495 CTGTGCAAGTGTGCGCGCGCGCAAGGCGCTGACCGACATCGTGGCGCTTGAACGAGGAG 1554  
 Db 1495 CTGTGCAAGTGTGCGCGCGCGCAAGGCGCTGACCGACATCGTGGCGCTTGAACGAGGAG 1554

1555 GCCGAGCTGAGCTGGCCGAGAACCGCGAGATCTCGCGAGCCCGTGCACGCGGTGTAC 1614  
 Db GCGGAGCTGAGCTGGCCGAGAACCGCGAGATCTCGCGAGCCCGTGCACGCGGTGTAC 1620  
 1615 TAGACCCCGAGCAAGCACTGTGTGCGCGAGATCCGAGAGCAGGGCCACGACAGTGGACC 1674  
 Db TAGACCCCGAGCAAGCACTGTGTGCGCGAGATCCGAGAGCAGGGCCACGACAGTGGACC 1680  
 1675 TACGAGATCTACGAGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCCAAGATGGC 1734  
 Db TACGAGATCTACGAGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCCAAGATGGC 1740  
 1735 ACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGACAGAGTCCCATGGAG 1794  
 Db ACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGACAGAGTCCCATGGAG 1800  
 1795 AGCATGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAGGAGACCTGG 1854  
 Db AGCATGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAGGAGACCTGG 1860  
 1855 GAGACCTGTGTGACCGACTACTTGGCAGGCGCACTGTGATCCCGAGTGGAGTTCTGTGAC 1914  
 Db GAGACCTGTGTGACCGACTACTTGGCAGGCGCACTGTGATCCCGAGTGGAGTTCTGTGAC 1920  
 1915 ACCCCCCCTGCTGAGCTGTGTGATACGAGCTGGAGAGGAGGCCCATCATCGGCGCGAG 1974  
 Db ACCCCCCCTGCTGAGCTGTGTGATACGAGCTGGAGAGGAGGCCCATCATCGGCGCGAG 1980  
 1975 ACCTTCTACGTGGAACCGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCTACGTG 2034  
 Db ACCTTCTACGTGGAACCGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCTACGTG 2040  
 2035 ACCGACCGGGCGCGGAGAGATCTGTAGCTTGAACGAGACCAACCGAGAGCCGAG 2094  
 Db ACCGACCGGGCGCGGAGAGATCTGTAGCTTGAACGAGACCAACCGAGAGCCGAG 2100  
 2095 CTGACAGGCCATCAGCTGGCCCTGAGGACGAGGAGGAGTGAACATCGTGACCGAC 2154  
 Db CTGACAGGCCATCAGCTGGCCCTGAGGACGAGGAGGAGTGAACATCGTGACCGAC 2160  
 2155 AGCCAGTAGCCCTGGGCATCATCCAGGCCCGCCGAGCAAGCGAGAGCGAGCTGGTG 2214  
 Db AGCCAGTAGCCCTGGGCATCATCCAGGCCCGCCGAGCAAGCGAGAGCGAGCTGGTG 2220  
 2215 ACCGAGATCATCAGAGCTGTATCAAGAGGAGAGGTGTACTGAGCTGGTGGCCGCGC 2274  
 Db ACCGAGATCATCAGAGCTGTATCAAGAGGAGAGGTGTACTGAGCTGGTGGCCGCGC 2280  
 2275 CACAAGGGCATCGGCGCAACGAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGCAAG 2334  
 Db CACAAGGGCATCGGCGCAACGAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGCAAG 2340  
 2335 GTGCTGTTCTTGGACCGGCAATGATGGCGCATCGTGTATCTACAGTATGAGCAGACCTG 2394  
 Db GTGCTGTTCTTGGACCGGCAATGATGGCGCATCGTGTATCTACAGTATGAGCAGACCTG 2400  
 2395 TACGTGGGAGCGGCGGCGCTTGGATCGATTAAAGCTTCCCGGGGCTAGCAACCGGTGA 2454  
 Db TACGTGGGAGCGGCGGCGCTTGGATCGATTAAAGCTTCCCGGGGCTAGCAACCGGTGA 2460  
 2455 TTC 2457  
 2461 TTC 2463

RESULT 6

US-10-190-435-10  
 ; Sequence 10, Application US/10190435  
 ; Publication No. US20030143248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEDEDE, Jan  
 ; APPLICANT: BARNETT, Susan W.  
 ; APPLICANT: LIAN, Ying

; APPLICANT: ENGELBRECHT, Susan  
 ; APPLICANT: VAN RENSBURG, Estrelita J.  
 ; TITLE OF INVENTION: POLYPEPTIDES ENCODING ANTIGENIC HIV TYPE C  
 ; FILE REFERENCE: PP18133.003 / 2302-18133  
 ; CURRENT APPLICATION NUMBER: US/10/190,435  
 ; CURRENT FILING DATE: 2002-12-30  
 ; NUMBER OF SEQ ID NOS: 319  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 10  
 ; LENGTH: 3930  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: GagCompPolmutAtt\_C  
 US-10-190-435-10

Query Match 99.1%; Score 2434.8; DB 15; Length 3930;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 14 TGGCCGAGGCATGAGCCAGGCCACAGGCCAAATCTGTATGAGCGCAGCAACTTCA 73  
 Db 1487 TCGCCGAGGCATGAGCCAGGCCACAGGCCAAATCTGTATGAGCGCAGCAACTTCA 1546  
 QY 74 AGGGCCCAAGGCATCATCAAGTCTTCAACTGCGCAAGAGGGCCACATCGGCCCA 133  
 Db 1547 AGGGCCCAAGGCATCATCAAGTCTTCAACTGCGCAAGAGGGCCACATCGGCCCA 1606  
 QY 134 ACTGCGCGCCCGCCGCAAGAGGCTGCTGGAAGTGGCAAGAGGGCCACACAGATGA 193  
 Db 1607 ACTGCGCGCCCGCCGCAAGAGGCTGCTGGAAGTGGCAAGAGGGCCACACAGATGA 1666  
 QY 194 AGGACTGCAACGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGTGCTTCCCGCAGGCA 253  
 Db 1667 AGGACTGCAACGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGTGCTTCCCGCAGGCA 1726  
 QY 254 AGGCGCGGAGTTCCTCCAGCGAGCAAGACCGGCCCAACAGCCCAACAGCGCGAGCTGC 313  
 Db 1727 AGGCGCGGAGTTCCTCCAGCGAGCAAGACCGGCCCAACAGCCCAACAGCGCGAGCTGC 1786  
 QY 314 AGGTGCGCGCGGCAAAACCCCGCGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 Db 1787 AGGTGCGCGCGGCAAAACCCCGCGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
 QY 374 TCCCCCAGATCACTGTGTGAGCGCCCTTGTGTGAGCATCAAGTGTGGCGCGCGCAGATCA 433  
 Db 1847 TCCCCCAGATCACTGTGTGAGCGCCCTTGTGTGAGCATCAAGTGTGGCGCGCGCAGATCA 1906  
 QY 434 AGGAGGCGCTGTGAGCACCGCGCGCGGACACCGTGTCTGAGGAGATGAGCTGCGCG 493  
 Db 1907 AGGAGGCGCTGTGAGCACCGCGCGCGGACACCGTGTCTGAGGAGATGAGCTGCGCG 1966  
 QY 494 GCAAGTGAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCAATCAAGTGTGCGCGCAGTACG 553  
 Db 1967 GCAAGTGAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCAATCAAGTGTGCGCGCAGTACG 2026  
 QY 554 ACCAGATCCTGTATCGAGATCTGCGGCAAGAGGCCATCGGCAACCTGTGTATCGGCGCCA 613  
 Db 2027 ACCAGATCCTGTATCGAGATCTGCGGCAAGAGGCCATCGGCAACCTGTGTATCGGCGCCA 2086  
 QY 614 CCCCCTGAAACATCATCGGCGCGCAACATGTCAGCCAGCTGGCTGGCGACCTGAACTTCC 673  
 Db 2087 CCCCCTGAAACATCATCGGCGCGCAACATGTCAGCCAGCTGGCTGGCGACCTGAACTTCC 2146  
 QY 674 CCATCAGCCCCCATCGAGACCGGTGCGCGTGAAGCTGAAGCTGAAGCTGAAGCTGAAGCTGAAG 733  
 Db 2147 CCATCAGCCCCCATCGAGACCGGTGCGCGTGAAGCTGAAGCTGAAGCTGAAGCTGAAGCTGAAG 2206  
 QY 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCCATCTCTGCGAGGAGA 793  
 Db 2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCCATCTCTGCGAGGAGA 2266

QY 794 TGGAGAGGCGGCAAGATCAACCAAGATCGGCGCGGAGAACCCCTTAAACACCCCGGTGT 853  
Db 2267 TGGAGAGGCGGCAAGATCAACCAAGATCGGCGCGGAGAACCCCTTAAACACCCCGGTGT 2326  
QY 854 TCGCCATCAAGAAGAGCAGCAGCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGA 913  
Db 2327 TCGCCATCAAGAAGAGCAGCAGCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGA 2386  
QY 914 ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGCTGGGCATCCCCACCCCGCGGCTGA 973  
Db 2387 ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGCTGGGCATCCCCACCCCGCGGCTGA 2446  
QY 974 AGAAGAGAGAGCGTGAACCGTGTGGAGGTGGGAGCGGCTACTTACGCGTGCCTCTGG 1033  
Db 2447 AGAAGAGAGAGCGTGAACCGTGTGGAGGTGGGAGCGGCTACTTACGCGTGCCTCTGG 2506  
QY 1034 ACGAGGACTTCCGCAAGTACACCGCTTACCATCTCCCGAGCATCAACAAACGAGACCCCG 1093  
Db 2507 ACGAGGACTTCCGCAAGTACACCGCTTACCATCTCCCGAGCATCAACAAACGAGACCCCG 2566  
QY 1094 GCATCCGCTACCAAGTACCAAGTGTGCTGCCCCAGGGCTGGAGGGCAGCCCCAGCATCTTCC 1153  
Db 2567 GCATCCGCTACCAAGTACCAAGTGTGCTGCCCCAGGGCTGGAGGGCAGCCCCAGCATCTTCC 2626  
QY 1154 AGAGCAGCATGACCAAGTACCAAGTGTGCTGCCCCAGGGCTGGAGGGCAGCCCCAGCATCTTCC 1213  
Db 2627 AGAGCAGCATGACCAAGTACCAAGTGTGCTGCCCCAGGGCTGGAGGGCAGCCCCAGCATCTTCC 2686  
QY 1214 ACCAGGCCCCCTGTAGCTGGGCGAGCATCTGGAGATCGGCGAGCAGCGCCCAAGATCG 1273  
Db 2687 ACCAGGCCCCCTGTAGCTGGGCGAGCATCTGGAGATCGGCGAGCAGCGCCCAAGATCG 2746  
QY 1274 AGGAGCTGCGCAAGCAGCATCTGCTGGGCTGGGCTTTCAACACCCCGCAAGAGCACCAGA 1333  
Db 2747 AGGAGCTGCGCAAGCAGCATCTGCTGGGCTGGGCTTTCAACACCCCGCAAGAGCACCAGA 2806  
QY 1334 AGGAGCCCCCTTCTGCGCATGAGCTGCAACCCCGCAAGTGAACCGTGCAGCCCATCG 1393  
Db 2807 AGGAGCCCCCTTCTGCGCATGAGCTGCAACCCCGCAAGTGAACCGTGCAGCCCATCG 2866  
QY 1394 AGCTGCCCGAGAGGAGCTGACCGTGAACGACATCCAGAACCTGGTGGCGAGCTGA 1453  
Db 2867 AGCTGCCCGAGAGGAGCTGACCGTGAACGACATCCAGAACCTGGTGGCGAGCTGA 2926  
QY 1454 ACTGGGCGAGCAGATCTACCCCGGCAATCAAGTGGCGGCTGTGAAGCTGTGGCGG 1513  
Db 2927 ACTGGGCGAGCAGATCTACCCCGGCAATCAAGTGGCGGCTGTGAAGCTGTGGCGG 2986  
QY 1514 GCGCAAGGCGCTGACCGACATCTGCTGCCCTGACCGAGGCGGAGCTGGAGCTGGCGG 1573  
Db 2987 GCGCAAGGCGCTGACCGACATCTGCTGCCCTGACCGAGGCGGAGCTGGAGCTGGCGG 3046  
QY 1574 AGAACCGGAGATCTGCGGAGCCCGTGCAGCGGCTGTACTAGACCCCGAGCAGAC 1633  
Db 3047 AGAACCGGAGATCTGCGGAGCCCGTGCAGCGGCTGTACTAGACCCCGAGCAGAC 3106  
QY 1634 TGGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTTACAGATCTTACAGAGC 1693  
Db 3107 TGGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTTACAGATCTTACAGAGC 3166  
QY 1694 CCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGCCGACACCAACGAGC 1753  
Db 3167 CCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGCCGACACCAACGAGC 3226  
QY 1754 TGAAGCAGCTGACCGGCGGCTGAGAGATCGCCATGGAGAGCATGTGATCTGGGGCA 1813  
Db 3227 TGAAGCAGCTGACCGGCGGCTGAGAGATCGCCATGGAGAGCATGTGATCTGGGGCA 3286  
QY 1814 AGACCCCAAGTCTCGGCTGCGCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACT 1873  
Db 3287 AGACCCCAAGTCTCGGCTGCGCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACT 3346  
QY 1874 ACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCTGTGAAGC 1933

Db 3347 ACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCTGTGAAGC 3406  
QY 1934 TGTGTTACCAAGTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGAGACGCG 1993  
Db 3407 TGTGTTACCAAGTGGAGAGGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGAGACGCG 3466  
QY 1994 CCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGCGCGGAGCA 2053  
Db 3467 CCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGCGCGGAGCA 3526  
QY 2054 AGATCGTGGCTGACCGAGACCAACCAACCAAGAGACCGAGCTGAGGCCCATTCAGCTGG 2113  
Db 3527 AGATCGTGGCTGACCGAGACCAACCAACCAAGAGACCGAGCTGAGGCCCATTCAGCTGG 3586  
QY 2114 CCTGTCAGGACGAGCGGCGAGCGAGGTGAACATCTGTGACCGAGCAGCTACGCCCTGGGCA 2173  
Db 3587 CCTGTCAGGACGAGCGGCGAGCGAGGTGAACATCTGTGACCGAGCAGCTACGCCCTGGGCA 3646  
QY 2174 TCATCCAGGCGCCAGCCCGACAGAGCGAGCGAGCTGGTGAACCAAGATCATTCAGCAGC 2233  
Db 3647 TCATCCAGGCGCCAGCCCGACAGAGCGAGCGAGCTGGTGAACCAAGATCATTCAGCAGC 3706  
QY 2234 TGATCAAGAGGAGAAAGTGTACCTGAGCTGGGTGCCCGCCCAAGGGGCAATCGCGGCA 2293  
Db 3707 TGATCAAGAGGAGAAAGTGTACCTGAGCTGGGTGCCCGCCCAAGGGGCAATCGCGGCA 3766  
QY 2294 ACAGAGCAGATCGACCAAGCTGGTGAAGGAGGATCCCGCAGGTGCTGTCTGAGACGCA 2353  
Db 3767 ACAGAGCAGATCGACCAAGCTGGTGAAGGAGGATCCCGCAGGTGCTGTCTGAGACGCA 3826  
QY 2354 TCAGTGGCGCATCGTGTATCTACAGTACATGAGCAGCTGTACGTGGGCGAGCGCGGCC 2413  
Db 3827 TCAGTGGCGCATCGTGTATCTACAGTACATGAGCAGCTGTACGTGGGCGAGCGCGGCC 3886  
QY 2414 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 3887 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3924

RESULT 7  
US-10-190-435-11  
; Sequence 11, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 3930  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna\_C  
US-10-190-435-11

Query Match 99.1%; Score 2434.8; DB 15; Length 3930;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 14 TGGCCGAGGCGCATGAGCCAGGCGCCACAGCGCCACATCTGTGATGCGAGCGGCAACTTCA 73  
Db 1487 TCGCCGAGGCGCATGAGCCAGGCGCCACAGCGCCACATCTGTGATGCGAGCGGCAACTTCA 1546

QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGGGCAAGAGGGCCACATCGCCGCA 133  
DB 1547 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGGGCAAGAGGGCCACATCGCCGCA 1606  
QY 134 ACTGGCGCCCCCGCAAGAAAGGGTGTGGAAGTGGCGCAAGAGGGCCACAGATGA 193  
DB 1607 ACTGGCGCCCCCGCAAGAAAGGGTGTGGAAGTGGCGCAAGAGGGCCACAGATGA 1666  
QY 194 AGGACTGCACGAGCGCCAGGCCAACTTCTTCGGGAGGACCTGGCTTCCGCCAGGGCA 253  
DB 1667 AGGACTGCACGAGCGCCAGGCCAACTTCTTCGGGAGGACCTGGCTTCCGCCAGGGCA 1726  
QY 254 AGGCGCGGAGTTCCCGACGAGCAGAACCGCGCAACAGCCCCACAGCGCGGAGCTGC 313  
DB 1727 AGGCGCGGAGTTCCCGACGAGCAGAACCGCGCAACAGCCCCACAGCGCGGAGCTGC 1786  
QY 314 AGGTGCGCGGCACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
DB 1787 AGGTGCGCGGCACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
QY 374 TCCCCAGATCAACCTGTGGGAGCGCCCTTGTGTAGCATCAAGGTGGCGGCGCGATCA 433  
DB 1847 TCCCCAGATCAACCTGTGGGAGCGCCCTTGTGTAGCATCAAGGTGGCGGCGCGATCA 1906  
QY 434 AGGAGCCCTGTGGACACCGCGCGCGAGCACCGTGTGGAGGAGTGAAGTGGCGCGCG 493  
DB 1907 AGGAGCCCTGTGGCACCGCGCGCGAGCACCGTGTGGAGGAGTGAAGTGGCGCGCG 1966  
QY 494 GCAAGTGAAGCCCAAGATGATCGCGGCGCATCGCGGCGTTCATCAAGGTGGCGCGATCG 553  
DB 1967 GCAAGTGAAGCCCAAGATGATCGCGGCGCATCGCGGCGTTCATCAAGGTGGCGCGATCG 2026  
QY 554 ACCAGATCTGTATGAGATCTCGCGGCAAGAGCGCATCGGCAACCGTGTGTATCGGCGCA 613  
DB 2027 ACCAGATCTGTATGAGATCTCGCGGCAAGAGCGCATCGGCAACCGTGTGTATCGGCGCA 2086  
QY 614 CCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCACTTGAATTC 673  
DB 2087 CCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCACTTGAATTC 2146  
QY 674 CCATCAGCCCATCGAGACGTCGCGGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGG 733  
DB 2147 CCATCAGCCCATCGAGACGTCGCGGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGG 2206  
QY 734 TGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGGAGGAGA 793  
DB 2207 TGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGGAGGAGA 2266  
QY 794 TGGAGAAGGAGGCAAGATCAACAAGATCGGCGCCGAGAACCCCTACAAACCCCGTGT 853  
DB 2267 TGGAGAAGGAGGCAAGATCAACAAGATCGGCGCCGAGAACCCCTACAAACCCCGTGT 2326  
QY 854 TCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGA 913  
DB 2327 TCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGA 2386  
QY 914 ACAAGCGCACCCAGACTTCTGGGAGGTGACGCTGGGCGATCCCCACCCCGCGGCTGA 973  
DB 2387 ACAAGCGCACCCAGACTTCTGGGAGGTGACGCTGGGCGATCCCCACCCCGCGGCTGA 2446  
QY 974 AGAAGAAGAGCGGTGACCGTGTGGAGTGGGCGAGCGCTTTCAGCGTGGCCCTGG 1033  
DB 2447 AGAAGAAGAGCGGTGACCGTGTGGAGTGGGCGAGCGCTTTCAGCGTGGCCCTGG 2506  
QY 1034 ACGAGACTTCCGCAAGTACACCGCCTTCAACATCCCGAGCATCAACACAGAGACCCCG 1093  
DB 2507 ACGAGACTTCCGCAAGTACACCGCCTTCAACATCCCGAGCATCAACACAGAGACCCCG 2566  
QY 1094 GCATCGCTACAGTACACGCTGTGCCCCAGGGCTGGAGGGAGCCCCAGCATCTTCC 1153  
DB 2567 GCATCGCTACAGTACACGCTGTGCCCCAGGGCTGGAGGGAGCCCCAGCATCTTCC 2626  
QY 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213

DB 2627 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCT 2686  
QY 1214 ACCAGGCCCCCTGTACCTGGGCGAGGACCTGGAGATCGGCGAGCACCGCGCAAGATCG 1273  
DB 2687 ACCAGGCCCCCTGTACCTGGGCGAGGACCTGGAGATCGGCGAGCACCGCGCAAGATCG 2746  
QY 1274 AGGAGCTGGCAAGCACCTGTGCGCTGGGGCTTACACACCCCGCAAGAGACCAAGCA 1333  
DB 2747 AGGAGCTGGCAAGCACCTGTGCGCTGGGGCTTACACACCCCGCAAGAGACCAAGCA 2806  
QY 1334 AGGAGCCCCCTTCTGCTCCCATCGAGCTGCACCCCGCAAGTGAACCGTGCAGCCCATCG 1393  
DB 2807 AGGAGCCCCCTTCTGCTCCCATCGAGCTGCACCCCGCAAGTGAACCGTGCAGCCCATCG 2866  
QY 1394 AGCTGCCGAGAGGAGAGCTGGACCGGTGAACGACATCAAGAGCTGGTGGGCAAGCTGA 1453  
DB 2867 AGCTGCCGAGAGGAGAGCTGGACCGGTGAACGACATCAAGAGCTGGTGGGCAAGCTGA 2926  
QY 1454 ACTGGGCGAGCCAGATCTACCCCGCATCAAGTGGCGGAGCTGTGAAGCTGTGCGCG 1513  
DB 2927 ACTGGGCGAGCCAGATCTACCCCGCATCAAGTGGCGGAGCTGTGAAGCTGTGCGCG 2986  
QY 1514 GCGCAAGGCCCTGACCGACATCGTGCCTTGAACCGAGGCGCGAGCTGGAGCTGGCGG 1573  
DB 2987 GCGCAAGGCCCTGACCGACATCGTGCCTTGAACCGAGGCGCGAGCTGGAGCTGGCGG 3046  
QY 1574 AGAACCGGAGATCTCTGCGGAGCCCGTGCACGGGTGTATCAACGCCAGCAAGGACC 1633  
DB 3047 AGAACCGGAGATCTCTGCGGAGCCCGTGCACGGGTGTATCAACGCCAGCAAGGACC 3106  
QY 1634 TGGTGGCGAGATCCAGAGCAGGCGCACGACCTGAGTGGACCTTACAGATCTTACCAAGGAGC 1693  
DB 3107 TGGTGGCGAGATCCAGAGCAGGCGCACGACCTGAGTGGACCTTACAGATCTTACCAAGGAGC 3166  
QY 1694 CTTTCAAGAACCTGAAGACCCGCAAGTACGCCAAGATCGCACCCGCCACACCAACGAGC 1753  
DB 3167 CTTTCAAGAACCTGNAAGACCCGCAAGTACGCCAAGATCGGCACCCGCCACCAACGAGC 3226  
QY 1754 TGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATCTGTGATCTGGGGA 1813  
DB 3227 TGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATCTGTGATCTGGGGA 3286  
QY 1814 AGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGGGAGACCTTGGTGGAGCCGACT 1873  
DB 3287 AGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTTGGGAGACCTTGGTGGAGCCGACT 3346  
QY 1874 ACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACAACCCCGCCCTTGGTGAAGC 1933  
DB 3347 ACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACAACCCCGCCCTTGGTGAAGC 3406  
QY 1934 TGTGTGTAACAGTGTGAAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGG 1993  
DB 3407 TGTGTGTAACAGTGTGAAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGG 3466  
QY 1994 CCGCAACCGCGAGACCAAGATCCGCAAGGCGGCTAGTGACCGACCGGGGCCGCGCA 2053  
DB 3467 CCGCAACCGCGAGACCAAGATCCGCAAGGCGGCTAGTGACCGACCGGGGCCGCGCA 3526  
QY 2054 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTGG 2113  
DB 3527 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTGG 3586  
QY 2114 CCCTGCAAGACAGCGGCAAGGAGTGAACATCGTGAACGAGCAGCTAAGCTGTGGGGA 2173  
DB 3587 CCCTGCAAGACAGCGGCAAGGAGTGAACATCGTGAACGAGCAGCTAAGCTGTGGGGA 3646  
QY 2174 TCATCCAGGCCAGCCCGACAGAGGAGCGAGCTGTGAACCAAGATCATCGAGCAGC 2233  
DB 3647 TCATCCAGGCCAGCCCGACAGAGGAGCGAGCTGTGAACCAAGATCATCGAGCAGC 3706  
QY 2234 TCATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGCGCGCCCAAGGGCATCTCGCGGCA 2293

Db 3707 TGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCCAACAGGGGCAATCGCGGCA 3766  
Qy 2294 ACAGAGAGATCGACAAGCTGGTGAAGCAAGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 2353  
Db 3767 ACAGAGAGATCGACAAGCTGGTGAAGCAAGGCATCCGCAAGGTGCTGTTCCTGGACGGCA 3826  
Qy 2354 TCGATGGCGGCATCGTGTATCTACCAAGTACATGAGGACCTGTAGTGGGAGCGCGGCC 2413  
Db 3827 TCGATGGCGGCATCGTGTATCTACCAAGTACATGAGGACCTGTAGTGGGAGCGCGGCC 3886  
Qy 2414 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 3887 CTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3924

RESULT 8  
US-10-190-435-58  
; Sequence 58, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBERG, Esrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 58  
; LENGTH: 5184  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefagCpolIna C  
US-10-190-435-58

Query Match 99.1%; Score 2434.8; DB 15; Length 5184;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCGCATGAGCGAGCCACAGCGCCATCTTGATCAGCGAGCAACTTCA 73  
Db 2741 TCGCCGAGGCGCATGAGCGAGCCACAGCGCCATCTTGATCAGCGAGCAACTTCA 2800

Qy 74 AGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCGCAATCGCCGCA 133  
Db 2801 AGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCGCAATCGCCGCA 2860

Qy 134 ACTGCGGCGCCCCCGCAAGAGGGCTGCTGGAGTGGCGAGGAGGCGCCACAGATGA 193  
Db 2861 ACTGCGGCGCCCCCGCAAGAGGGCTGCTGGAGTGGCGAGGAGGCGCCACAGATGA 2920

Qy 194 AGGACTGCACCGAGCGCCAGCCAACTTCTTCCGCGAGGACCTGCGCTTCCCGCAGGGCA 253  
Db 2921 AGGACTGCACCGAGCGCCAGCCAACTTCTTCCGCGAGGACCTGCGCTTCCCGCAGGGCA 2980

Qy 254 AGGCCCCGAGATTCCCGAGGAGCAGAAACCGCGCCAAACAGCCCCACAGCGGAGCTGC 313  
Db 2981 AGGCCCCGAGATTCCCGAGGAGCAGAAACCGCGCCAAACAGCCCCACAGCGGAGCTGC 3040

Qy 314 AGGTGCGGCGGCAACCCCCGAGCGAGGCGCGCGCGCGAGCGCCAGGCAACCTGAACT 373  
Db 3041 AGGTGCGGCGGCAACCCCCGAGCGAGGCGCGCGCGCGAGCGCCAGGCAACCTGAACT 3100

Qy 374 TCCCCCAGATCACCTCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGCGGCGCAGATCA 433  
Db 3101 TCCCCCAGATCACCTCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGCGGCGCAGATCA 3160

Qy 434 AGGAGGCCCTGCTGGACACCGGCGCGAGCACACCGTGTGGAGGAGATGAGCTGCGCG 493

Db 3161 AGGAGGCCCTGCTGGCCACCGCGCGCCGACGACACCTGCTGGAGGAGATGAGCTGCGCG 3220  
Qy 494 GCAAGTGGAAAGCCCAAGATGATCGGCGGCATTCGCGCGCTTTCATCAAGGTGGCGCAGTACG 553  
Db 3221 GCAAGTGGAAAGCCCAAGATGATCGGCGGCATTCGCGCGCTTTCATCAAGGTGGCGCAGTACG 3280  
Qy 554 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCATTCGCGCACCGTGTGATCGGCCCCCA 613  
Db 3281 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCATTCGCGCACCGTGTGATCGGCCCCCA 3340  
Qy 614 CCCCCGTGAACATCATCGGCGGCAACATGTCGACCCAGCTGGGGTGCACCTCTGAACCTTCC 673  
Db 3341 CCCCCGTGAACATCATCGGCGGCAACATGTCGACCCAGCTGGGGTGCACCTCTGAACCTTCC 3400  
Qy 674 CCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGTAAGCCCGGCATCGGACGCCCCCAAGG 733  
Db 3401 CCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGTAAGCCCGGCATCGGACGCCCCCAAGG 3460  
Qy 734 TGAAGCAGTGGCCCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793  
Db 3461 TGAAGCAGTGGCCCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 3520  
Qy 794 TGAAGAGGAGGCGCAAGATCAACAGATCGGCGCCGAGAACCCCTACAACACCCCGTGT 853  
Db 3521 TGAAGAGGAGGCGCAAGATCAACAGATCGGCGCCGAGAACCCCTACAACACCCCGTGT 3580  
Qy 854 TCGCCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 913  
Db 3581 TCGCCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3640  
Qy 914 ACAAGCGCACCCAGGACTTCTTGGGAGGTGACGCTGGGCATCCCCACACCCCGCGCCCTGA 973  
Db 3641 ACAAGCGCACCCAGGACTTCTTGGGAGGTGACGCTGGGCATCCCCACACCCCGCGCCCTGA 3700  
Qy 974 AGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1033  
Db 3701 AGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3760  
Qy 1034 ACCAGGAGCTTCCGCAAGTACACCGCTTCCACCATCCCGACGATCCCGACACCCCGCGCCCG 1093  
Db 3761 ACCAGGAGCTTCCGCAAGTACACCGCTTCCACCATCCCGACGATCCCGACACCCCGCGCCCG 3820  
Qy 1094 GCATCCGCTACCAAGTACCAAGTGTGCGCCAGAGGCTGGAAGGGCAGCGCCAGCATCTTCC 1153  
Db 3821 GCATCCGCTACCAAGTACCAAGTGTGCGCCAGAGGCTGGAAGGGCAGCGCCAGCATCTTCC 3880  
Qy 1154 AGAGCAGCATGACCAAGATCTTGGAGGCTTTCGCGCGCCGCAACCCCGAGATCGTGTCT 1213  
Db 3881 AGAGCAGCATGACCAAGATCTTGGAGGCTTTCGCGCGCCGCAACCCCGAGATCGTGTCTCT 3940  
Qy 1214 ACCAGGCCCGCTGTAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1273  
Db 3941 ACCAGGCCCGCTGTAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4000  
Qy 1274 AGAGGCTGCGCAAGACCTGCTGCTGGGCTTTCACACCCCGCAACAGAGCAGCAGCA 1333  
Db 4001 AGAGGCTGCGCAAGACCTGCTGCTGGGCTTTCACACCCCGCAACAGAGCAGCAGCA 4060  
Qy 1334 AGAGGCCCGCTTCTGCGCCATCGAGTGCACCCCGCAACAGTGGAGCGTGCAGGCCATTCG 1393  
Db 4061 AGAGGCCCGCTTCTGCGCCATCGAGTGCACCCCGCAACAGTGGAGCGTGCAGGCCATTCG 4120  
Qy 1394 AGCTGCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1453  
Db 4121 AGCTGCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4180  
Qy 1454 ACTGGGCGCAGCAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 1513  
Db 4181 ACTGGGCGCAGCAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGCTGCTGCGCG 4240  
Qy 1514 GCGCCAAGGCCCTGACCGACATCGTGCCTTGCACCGAGGAGGCGAGCTGAGCTGCGCG 1573



```

Db 4241 GCGCAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAGCCGAGCTGGAGCTGGCGG 4300
Qy 1574 AGAACCGGAGATCTTGGCGGAGCCCGTGCACGGGTGTACTACGACCCAGCAAGGACC 1633
Db 4301 AGAACCGGAGATCTTGGCGGAGCCCGTGCACGGGTGTACTACGACCCAGCAAGGACC 4360
Qy 1634 TGGTGGCCGAGATCCAGAAAGCAGGCGCCAGCACAGTGGACCTTACAGATCTTACCAAGGAGC 1693
Db 4361 TGGTGGCCGAGATCCAGAAAGCAGGCGCCAGCACAGTGGACCTTACAGATCTTACCAAGGAGC 4420
Qy 1694 CTTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATGCGCACCCGCCACACCAACGAGC 1753
Db 4421 CTTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATGCGCACCCGCCACACCAACGAGC 4480
Qy 1754 TGAAGCAGCTGACCGAGGCGGTGCGAGAGATGCCATGAGAGGATCTGTGATCTGGGGCA 1813
Db 4481 TGAAGCAGCTGACCGAGGCGGTGCGAGAGATGCCATGAGAGGATCTGTGATCTGGGGCA 4540
Qy 1814 AGACCCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGACCGACT 1873
Db 4541 AGACCCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGACCGACT 4600
Qy 1874 ACTGCAGGCCACCTGGATCCCGAGTGGAGTTTCGTGAACACCCCGCCCTTGGTGAAGC 1933
Db 4601 ACTGCAGGCCACCTGGATCCCGAGTGGAGTTTCGTGAACACCCCGCCCTTGGTGAAGC 4660
Qy 1934 TGTGTACAGCTGAGAGAGGAGCCATCATCGGCGCGGAGACCTTCTACGTGGAGCGGG 1993
Db 4661 TGTGTACAGCTGAGAGAGGAGCCATCATCGGCGCGGAGACCTTCTACGTGGAGCGGG 4720
Qy 1994 CGGCAACCGCGAGACCAAGATCGGCAAGGCGCGGTACGTGACGACCGGCGCGGAGAG 2053
Db 4721 CGGCAACCGCGAGACCAAGATCGGCAAGGCGCGGTACGTGACGACCGGCGCGGAGAG 4780
Qy 2054 AGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGG 2113
Db 4781 AGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGG 4840
Qy 2114 CCTTCAGAGACCGGCGAGCGAGGTGAACATCGTGAACGACGAGTACGCGCTTGGGCA 2173
Db 4841 CCTTCAGAGACCGGCGAGCGAGGTGAACATCGTGAACGACGAGTACGCGCTTGGGCA 4900
Qy 2174 TCATCCAGGCCAGCCGAGCAAGAGCGAGAGCGAGCTGTGTAACCAAGATCATCGAGGAGC 2233
Db 4901 TCATCCAGGCCAGCCGAGCAAGAGCGAGAGCGAGCTGTGTAACCAAGATCATCGAGGAGC 4960
Qy 2234 TGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCCCAAGGGCGCATCGCGGCA 2293
Db 4961 TGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCCCAAGGGCGCATCGCGGCA 5020
Qy 2294 AGAGCAGATCGACAAGCTGTGAGCAAGGGCATCGGCAAGGTGCTTCTTGGAGCGCA 2353
Db 5021 AGAGCAGATCGACAAGCTGTGAGCAAGGGCATCGGCAAGGTGCTTCTTGGAGCGCA 5080
Qy 2354 TCGATGGGGCATCTGTATCTACAGTATACGAGACCTGTACCTGGGCGAGCGCGGCC 2413
Db 5081 TCGATGGGGCATCTGTATCTACAGTATACGAGACCTGTACCTGGGCGAGCGCGGCC 5140
Qy 2414 CTAGGATCGATTAAGGCTTCCGGGGCTAGCACCGGT 2451
Db 5141 CTAGGATCGATTAAGGCTTCCGGGGCTAGCACCGGT 5178

```

RESULT 9

```

; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

```

```

; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190.305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 5184
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: TatRevNefgagCpolaIna C
US-10-190-305A-82

Query Match 99.1%; Score 2434.8; DB 16; Length 5184;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCCATGAGCCAGGCCACCGAGCCCAACATCTGTATGCGAGCGCAACTTCA 73
Db 2741 TGGCCGAGGCCATGAGCCAGGCCACCGAGCCCAACATCTGTATGCGAGCGCAACTTCA 2800
Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGGAGGGCCACATTCGCCCGCA 133
Db 2801 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGGAGGGCCACATTCGCCCGCA 2860
Qy 134 ACTGCGCGCGCCCGCGCAAGAGGGCTCTTGAAGTGGGCAAGGAGGGCCACCGAGTGA 193
Db 2861 ACTGCGCGCGCCCGCGCAAGAGGGCTCTTGAAGTGGGCAAGGAGGGCCACCGAGTGA 2920
Qy 194 AGGACTGACACGAGCGCGCCAGGCCCAACTTCTTCCGAGGAGCTGTGCCCTTCCCGCAGGGCA 253
Db 2921 AGGACTGACACGAGCGCGCCAGGCCCAACTTCTTCCGAGGAGCTGTGCCCTTCCCGCAGGGCA 2980
Qy 254 AGGGCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCCAAACAGCCCAACAGCGCGAGCTGC 313
Db 2981 AGGGCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCCAAACAGCGCGAGCTGC 3040
Qy 314 AGGTGCGCGGCGCAACACCCCGAGCGAGGCGGGCGCGGCGCGAGGGCACCTTGAACCT 373
Db 3041 AGGTGCGCGGCGCAACACCCCGAGCGAGGCGGGCGCGGCGCGAGGGCACCTTGAACCT 3100
Qy 374 TCCGCCAGATCACCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGCGCGCCAGATCA 433
Db 3101 TCCGCCAGATCACCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGCGCGCCAGATCA 3160
Qy 434 AGGAGGCCCTCTGTGACACCGCGCGCGACGACACCGTGTCTGAGAGAGATGAGCTTGC 493
Db 3161 AGGAGGCCCTCTGTGACACCGCGCGCGACGACACCGTGTCTGAGAGAGATGAGCTTGC 3220
Qy 494 GCAAGTGAAGGCCCAAGATGATCGGCGCATCGGCGGCTTTCATCAAGGTGGCGCGAGTAC 553
Db 3221 GCAAGTGAAGGCCCAAGATGATCGGCGCATCGGCGGCTTTCATCAAGGTGGCGCGAGTAC 3280
Qy 554 ACCAGATCTGTATCGAGATCTTGGGCAAGAGGCCATCGGACCGTGTGTGATCGGCCCA 613
Db 3281 ACCAGATCTGTATCGAGATCTTGGGCAAGAGGCCATCGGACCGTGTGTGATCGGCCCA 3340
Qy 614 CCCCCGTGAACATCATTCGCGCGCCCAACATGTCTGACCTGGGTGCAACCTTGAACCTTCC 673
Db 3341 CCCCCGTGAACATCATTCGCGCGCCCAACATGTCTGACCTGGGTGCAACCTTGAACCTTCC 3400
Qy 674 CCATCAGCCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGG 733
Db 3401 CCATCAGCCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGG 3460
Qy 734 TGAAGCAGTGGCCCTCTGACCGAGGAGAGATCAAGGCCCTTGAACCGCTTTCGCGAGGAGA 793
Db 3461 TGAAGCAGTGGCCCTCTGACCGAGGAGAGATCAAGGCCCTTGAACCGCTTTCGCGAGGAGA 3520
Qy 794 TGGAGAGGAGGGGCAAGATCAACCAAGATCGCGCCCGGAGAACCCCTTACAACACCCCGGT 853
Db 3521 TGGAGAGGAGGGGCAAGATCAACCAAGATCGCGCCCGGAGAACCCCTTACAACACCCCGGT 3580

```





Db 2281 |CACAGGGCATCGCGGCAACGACGATGACAAAGCTGGTGAACAAGGGCATCCGCAAG 2340  
Qy 2335 |GTGCTGTTCTCGGACGGCATCGATGGCGCATCGTGTATCTACCAAGTACATGACGACCTG 2394  
Db 2341 |GTGCTGTTCTCGGACGGCATCGATGGCGCATCGTGTATCTACCAAGTACATGACGACCTG 2400  
Qy 2395 |TACGTGGCAGCGCGGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 2401 |TACGTGGCAGCGCGGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

## RESULT 11

US-10-190-305A-38  
; Sequence 38, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MESEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190.305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM\_C  
US-10-190-305A-38

Query Match 98.8%; Score 2428.6; DB 16; Length 2457;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 1 |GTCACGCCACCATGCGCGGCGCATGAGCCAGCCGACAGCGCCCAACATCCTGATCGAG 60  
Db 1 |GTCACGCCACCATGCGCGGCGCATGAGCCAGCCGACAGCGCCCAACATCCTGATCGAG 60  
Qy 61 |CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 120  
Db 61 |CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 120  
Qy 121 |CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 |CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Qy 181 |GGCCACCAAGATGAGGAGTGCACAGGCGCGCGGCGCAACTTCTTCGCGGAGGACTGGCC 240  
Db 181 |GGCCACCAAGATGAGGAGTGCACAGGCGCGCGGCGCAACTTCTTCGCGGAGGACTGGCC 240  
Qy 241 |TTCCCGCCAGGCAAGGCGCGGAGTTCGCCAGCAGCAGAACCGCGCCCAACAGCCCCACC 300  
Db 241 |TTCCCGCCAGGCAAGGCGCGGAGTTCGCCAGCAGCAGAACCGCGCCCAACAGCCCCACC 300  
Qy 301 |AGCCGCGAGTGCAGGTGCGCGCGCAAAACCCCGCAGCAGGCGCGCGCGCGAGCGCCAG 360  
Db 301 |AGCCGCGAGTGCAGGTGCGCGCGCAAAACCCCGCAGCAGGCGCGCGCGCGAGCGCCAG 360  
Qy 361 |GGCACCTGACTTCCCGCCAGATCACCTGTGGCAGCGCCCTGTGGAGCATCAAGGTG 420  
Db 361 |GGCACCTGACTTCCCGCCAGATCACCTGTGGCAGCGCCCTGTGGAGCATCAAGGTG 420  
Qy 421 |GGCGGCGAGATCAAGGAGGCGCTGCTGACACCGCGCGCAGCAGACCGTGTCTGGAGGAG 480  
Db 421 |GGCGGCGAGATCAAGGAGGCGCTGCTGACACCGCGCGCAGCAGACCGTGTCTGGAGGAG 480  
Qy 481 |ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG 540  
Db 481 |ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG 540

Db 481 |ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG 540  
Qy 541 |GTGCGCCAGTACGACCAAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
Db 541 |GTGCGCCAGTACGACCAAGATCCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
Qy 601 |CTGATCGCGCCCAACCCCGGTGAACATCATCGGCGCAACATGTGTGACCCAGCTGGGTGC 660  
Db 601 |CTGATCGCGCCCAACCCCGGTGAACATCATCGGCGCAACATGTGTGACCCAGCTGGGTGC 660  
Qy 661 |ACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCGATG 720  
Db 661 |ACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCGATG 720  
Qy 721 |GACGGCCCAAGGTGAAGCAGTGGCCCTCTGACCGAGAGAGATCAAGAGCCCTCACCGCC 780  
Db 721 |GACGGCCCAAGGTGAAGCAGTGGCCCTCTGACCGAGAGAGATCAAGAGCCCTCACCGCC 780  
Qy 781 |ATCTGCGAGGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCCCCCGAGAACCCCTAC 840  
Db 781 |ATCTGCGAGGATGGAAGAGGAGGCAAGATCAACCAAGATCGGCCCCCGAGAACCCCTAC 840  
Qy 841 |AACACCCCGCTGTTCGCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
Db 841 |AACACCCCGCTGTTCGCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
Qy 901 |TTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGACGCTGGGCGATCCCCCAG 960  
Db 901 |TTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGACGCTGGGCGATCCCCCAG 960  
Qy 961 |CCCGCGGCTGAAGAGAGAGCGTGAACCGTGTGAGCGTGGGCGAGCGCTACTTTC 1020  
Db 961 |CCCGCGGCTGAAGAGAGAGCGTGAACCGTGTGAGCGTGGGCGAGCGCTACTTTC 1020  
Qy 1021 |AGCGTGGCTTGGACGAGGAGTTCGCGAAGTACACCGCTTCCACATCCCCCAGCATCAAC 1080  
Db 1021 |AGCGTGGCTTGGACGAGGAGTTCGCGAAGTACACCGCTTCCACATCCCCCAGCATCAAC 1080  
Qy 1081 |AACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCGCCCAAGGCTGGAAGGGGAGC 1140  
Db 1081 |AACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCGCCCAAGGCTGGAAGGGGAGC 1140  
Qy 1141 |CCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGGCGCTTCCGCGCGCCGCAACCCC 1200  
Db 1141 |CCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGGCGCTTCCGCGCGCCGCAACCCC 1200  
Qy 1201 |GAGATCGTGTATCTACAGGCGCCCTGTGTAGTGGGCGAGGAGCTGGAGATGCGGCGAGC 1260  
Db 1201 |GAGATCGTGTATCTACAGGCGCCCTGTGTAGTGGGCGAGGAGCTGGAGATGCGGCGAGC 1260  
Qy 1261 |CGCGCCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGCGCTTACCAACCCCGAG 1320  
Db 1261 |CGCGCCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGCGCTTACCAACCCCGAG 1320  
Qy 1321 |AAGAAGCACCAGAGGAGCGCCCTTCTCTGCGCATGAGGCTTACGAGCTGCAACCCGCAAG 1374  
Db 1321 |AAGAAGCACCAGAGGAGCGCCCTTCTCTGCGCATGAGGCTTACGAGCTGCAACCCGCAAG 1374  
Qy 1375 |TGAGCCGTGAGCCCATCGAGCTGCGCGAGAGGAGGAGCTGGACCGTGAACGAGCATCCAG 1434  
Db 1381 |TGAGCCGTGAGCCCATCGAGCTGCGCGAGAGGAGGAGCTGGACCGTGAACGAGCATCCAG 1440  
Qy 1435 |AAGCTGTGGGCAAGCTGAACCTGGCGCAGCAGATCTACCCCGGCGATCAAGGTGCGCGAG 1494  
Db 1441 |AAGCTGTGGGCAAGCTGAACCTGGCGCAGCAGATCTACCCCGGCGATCAAGGTGCGCGAG 1500  
Qy 1495 |CTGTGAAGTGTCTGGCGGCGCGCAAGCCCTTGAACGAGCATCTGTGCCCTTGAACGAGGAG 1554  
Db 1501 |CTGTGAAGTGTCTGGCGGCGCGCAAGCCCTTGAACGAGCATCTGTGCCCTTGAACGAGGAG 1560  
Qy 1555 |GCCGAGCTGAGCTGGCGCGAGAACCCGCGAGATCTGCGCGAGGCGCGTGCACGGGCTGTAC 1614  
Db 1561 |GCCGAGCTGAGCTGGCGCGAGAACCCGCGAGATCTGCGCGAGGCGCGTGCACGGGCTGTAC 1620

```

1615 TACGACCCAGCAAGGACTGTGGCCGAGATCCAGAGCAGGGCCAGCAAGTGGACC 1674
1621 TACGACCCAGCAAGGACTGTGGCCGAGATCCAGAGCAGGGCCAGCAAGTGGACC 1680
1675 TACGACATCTACGAGGAGCCCTTCAAGAACCTTGAAGACGGCAAGTACGCCAAGATGGC 1734
1681 TACGACATCTACGAGGAGCCCTTCAAGAACCTTGAAGACGGCAAGTACGCCAAGATGGC 1740
1735 ACCGCCCCACCAACGAGCTGAAGCAGTGAACGAGGCGGTGCAAGAGATGCCATGAG 1794
1741 ACCGCCCCACCAACGAGCTGAAGCAGTGAACGAGGCGGTGCAAGAGATGCCATGAG 1800
1795 AGCATCGTGTATCGGGGAGAGCCGCCAAGTTCGCTGCCCTGCCATCCAGAGGAGACCTGG 1854
1801 AGCATCGTGTATCGGGGAGAGCCGCCAAGTTCGCTGCCCTGCCATCCAGAGGAGACCTGG 1860
1855 GAGACCTGTGTGACCGCATCTACTGGCAGGCCACTTGGATCCCCGAGTGGGAGTTCTGTGAAC 1914
1861 GAGACCTGTGTGACCGCATCTACTGGCAGGCCACTTGGATCCCCGAGTGGGAGTTCTGTGAAC 1920
1915 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAAAGAGCCCATCATCGGCGCCGAG 1974
1921 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAAAGAGCCCATCATCGGCGCCGAG 1980
1975 ACCTTCTACGTGGAGCGGCCGCCAACCAGCGAGACCAAGATCGGCAAGCGGCTACGTG 2034
1981 ACCTTCTACGTGGAGCGGCCGCCAACCAGCGAGACCAAGATCGGCAAGCGGCTACGTG 2040
2035 ACCGACCGGGGCGGCGAGAAAGATCTGTGAGCCTTGACCGAGACCAACCAAGAGACCGAG 2094
2041 ACCGACCGGGGCGGCGAGAAAGATCTGTGAGCCTTGACCGAGACCAACCAAGAGACCGAG 2100
2095 GTGAGGCGCATCAGCTGGCCCTGTGAGCAGCGAGCGAGAGGTGAACATCTGTGACCGAC 2154
2101 CTGAGGCGCATCAGCTGGCCCTGTGAGCAGCGAGCGAGAGGTGAACATCTGTGACCGAC 2160
2155 AGCCAGTAGCCCTGGGCGATCATCAGCGCCAGCGCCGAGAGCGAGCGAGCGAGCTGGT 2214
2161 AGCCAGTAGCCCTGGGCGATCATCAGCGCCAGCGCCGAGAGCGAGCGAGCGAGCTGGT 2220
2215 AACCAAGATCATCAGCAGCTGATCAAGAGAGAGAGGTGTACTGTGGTGGTGGCCGCC 2274
2221 AACCAAGATCATCAGCAGCTGATCAAGAGAGAGAGGTGTACTGTGGTGGTGGCCGCC 2280
2275 CACAGAGGATCGGCGGCAACGAGCAGATCGAAGAGTGTGAGCAAGGGGATCCGCAAG 2334
2281 CACAGAGGATCGGCGGCAACGAGCAGATCGAAGAGTGTGAGCAAGGGGATCCGCAAG 2340
2335 GTGCTGTTCTCGAGCGGATCGATCGGCGCATCGTGTATCTACAGTACATGAGCAGCCTG 2394
2341 GTGCTGTTCTCGAGCGGATCGATCGGCGCATCGTGTATCTACAGTACATGAGCAGCCTG 2400
2395 TACGTGGGCGAGCGGGCGGCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451
2401 TACGTGGGCGAGCGGGCGGCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

```

RESULT 12

```

US-09-899-575-30
; Sequence 30, Application US/09899575
; Publication No. US2003023961A1
; GENERAL INFORMATION:
; APPLICANT: Zur Megede, Jan
; APPLICANT: Barnett, Susan W.
; APPLICANT: Egnelbrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janse
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP01631.102
; CURRENT APPLICATION NUMBER: US/09/899,575
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475,704

```

```

; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PR975(+)
US-09-899-575-30

```

Query Match 98.3%; Score 2415.4; DB 10; Length 2469;

Best Local Similarity 99.3%; Pred. No. 0;

Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

```

QY 1 GTGACGCGCACATGCGCGAGGCCATGAGCGCCACAGCGCCACAGCGCCAAATCTGTATGAG 60
DB 1 GTCACGCGCACATGCGCGAGGCCATGAGCGCCACAGCGCCACAGCGCCAAATCTGTATGAG 60
QY 61 CGCAGCAACTTCAAGGGCCCCAAGGCGCATCATCAAGTCTTCACTGCGGCAAGGAGGCG 120
DB 61 CGCAGCAACTTCAAGGGCCCCAAGGCGCATCATCAAGTCTTCACTGCGGCAAGGAGGCG 120
QY 121 CACATCGCCCCCAACTGCGCGCCCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180
DB 121 CACATCGCCCCCAACTGCGCGCCCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180
QY 181 GGCACACAGATGAAGGACTGCAACGAGCGCCAGCGCAAACTTCTTCGCGAGGACCTGGCC 240
DB 181 GGCACACAGATGAAGGACTGCAACGAGCGCCAGCGCAAACTTCTTCGCGAGGACCTGGCC 240
QY 241 TTCCCGCAGGCGCAAGGGCCCCGAGTTCCCGCAGCGAGAGAACCGCGCCAAACAGCCCCCACC 300
DB 241 TTCCCGCAGGCGCAAGGGCCCCGAGTTCCCGCAGCGAGAGAACCGCGCCAAACAGCCCCCACC 300
QY 301 AGCCGCGAGCTGCGAGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAG 360
DB 301 AGCCGCGAGCTGCGAGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAG 360
QY 361 GGCACCTGTAACTTCCCGCAGATCACTGTGGCGAGCGCCCGCTTGTGAGCATCAAGGTG 420
DB 361 GGCACCTGTAACTTCCCGCAGATCACTGTGGCGAGCGCCCGCTTGTGAGCATCAAGGTG 420
QY 421 GGCAGCAGATCAAGGAGGCGCTGCTGGACACCGCGCGCGACACACCGTGTCTGAGGAG 480
DB 421 GGCAGCAGATCAAGGAGGCGCTGCTGGACACCGCGCGCGACACACCGTGTCTGAGGAG 480
QY 481 ATGAGCTTGCCTCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGGCTTTCATCAAG 540
DB 481 ATGAGCTTGCCTCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGGCTTTCATCAAG 540
QY 541 GTGCGCGAGTACGACAGATCTGTATCGAGATCTGCGGCGAAGAGGCCATCGGCGACCGTG 600
DB 541 GTGCGCGAGTACGACAGATCTGTATCGAGATCTGCGGCGAAGAGGCCATCGGCGACCGTG 600
QY 601 CTGATCGGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
DB 601 CTGATCGGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
QY 661 ACCCTGAACTTCCCGCATCAGCGCCCATCGAGACCGTGGCCCGTGAAGCTGAAAGCCCGGCATG 720
DB 661 ACCCTGAACTTCCCGCATCAGCGCCCATCGAGACCGTGGCCCGTGAAGCTGAAAGCCCGGCATG 720
QY 721 GACGCGCCCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780
DB 721 GACGCGCCCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780
QY 781 ATCTGCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCGCCCGAGAACCCCTAC 840
DB 781 ATCTGCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCGCCCGAGAACCCCTAC 840
QY 841 AACACCCCGGTGTTCCGCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGGAC 900
DB 841 AACACCCCGGTGTTCCGCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGGAC 900

```





Db 1148 AGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGGCGCAAGAGGGGCCACATCGCCGCA 1207  
Qy 134 ACTGCGCGCCCCCGCAGAGAGGGCTGCTGAAGTGGCGCAAGAGGGGCCACAGATGA 193  
Db 1208 ACTGCGCGCCCCCGCAGAGAGGGCTGCTGAAGTGGCGCAAGAGGGGCCACAGATGA 1267  
Qy 194 AGGAATGCAACGAGCGCCAGGCGCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGCA 253  
Db 1268 AGGAATGCAACGAGCGCCAGGCGCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGCA 1327  
Qy 254 AGGCGCGGAGTTCCTCCAGCAGAGCAAAACCGCGCAACAGCCCAACAGCGCGCGAGCTGC 313  
Db 1328 AGGCGCGGAGTTCCTCCAGCAGAGCAAAACCGCGCAACAGCCCAACAGCGCGCGAGCTGC 1387  
Qy 314 AGGTGCGGCGCAACACCCCGCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
Db 1388 AGGTGCGGCGCAACACCCCGCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1447  
Qy 374 TCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGCGGCGCAGATCA 433  
Db 1448 TCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGCGGCGCAGATCA 1507  
Qy 434 AGGAGCCCTGTGCAACACCGCGCGCGACACACCGTGTGAGAGATGAGCCTGCCCG 493  
Db 1508 AGGAGGCCCTGTGCAACACCGCGCGCGACACACCGTGTGAGAGATGAGCCTGCCCG 1567  
Qy 494 GCAATGGAAGCCAGATGATCGGGGCGCATCGCGGCTTTCATCAAGGTGGCGCGAGTAG 553  
Db 1568 GCAATGGAAGCCAGATGATCGGGGCGCATCGCGGCTTTCATCAAGGTGGCGCGAGTAG 1627  
Qy 554 ACCAGATCCTGATCGAGATCTCGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 613  
Db 1628 ACCAGATCCTGATCGAGATCTCGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 1687  
Qy 614 CCCCCGTGAACATCATCGSCCGCAACATGCTGACCGAGCTGGGCTGCACCCCTGAATTC 673  
Db 1688 CCCCCGTGAACATCATCGSCCGCAACATGCTGACCGAGCTGGGCTGCACCCCTGAATTC 1747  
Qy 674 CCATCAGCCCCATCAGACCGTGGCGGTGAAGCTGAAGCCCGGATGAGCGGCCCAAGG 733  
Db 1748 CCATCAGCCCCATCAGACCGTGGCGGTGAAGCTGAAGCCCGGATGAGCGGCCCAAGG 1807  
Qy 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGAGGAGA 793  
Db 1808 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGAGGAGA 1867  
Qy 794 TGAAGAGGAGGCAAGATCAACAGATCGGCCCGAGAACCCCTACAAACCCCGGTGT 853  
Db 1868 TGAAGAGGAGGCAAGATCAACAGATCGGCCCGAGAACCCCTACAAACCCCGGTGT 1927  
Qy 854 TCGCCATCAAGAAGAGCAGACCAAGTGGCGCAAGCTGTGACCTTCGCGAGCTGA 913  
Db 1928 TCGCCATCAAGAAGAGGAGCAGCAACCAAGTGGCGCAAGCTGTGACCTTCGCGAGCTGA 1987  
Qy 914 ACAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGCATCCCCACCCCGCGCCTGA 973  
Db 1988 ACAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGCATCCCCACCCCGCGCCTGA 2047  
Qy 974 AGAAGAAGAAGACGTGACCGTGTGAGCTGGGCGAGCGCTTACAGCTGCCCCCTGG 1033  
Db 2048 AGAAGAAGAAGACGTGACCGTGTGAGCTGGGCGAGCGCTTACAGCTGCCCCCTGG 2107  
Qy 1034 ACGAGACTTCGCAAGTACACGGCTTCAACATCCCGAGCATCAACACGAGACCCCG 1093  
Db 2108 ACGAGACTTCGCAAGTACACGGCTTCAACATCCCGAGCATCAACACGAGACCCCG 2167  
Qy 1094 GCATCCGCTACAGTACACGTGTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTCC 1153  
Db 2168 GCATCCGCTACAGTACACGTGTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTCC 2227  
Qy 1154 AGACAGCATGACCAAGATCTCGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCT 1213  
Db 2228 AGACAGCATGACCAAGATCTCGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCT 2287

Qy 1214 ACCAGGCCCTGTGTACGTGGGCGAGCACTGGAGATCGGCAGCACCGCGCAAGATCG 1273  
Db 2288 ACCAGGCCCTGTGTACGTGGGCGAGCACTGGAGATCGGCAGCACCGCGCAAGATCG 2347  
Qy 1274 AGGAGCTGCGCAAGCACCTGTGCGCTGGGCTTCAACACCCCGCAAGAAGACCAAGA 1333  
Db 2348 AGGAGCTGCGCAAGCACCTGTGCGCTGGGCTTCAACACCCCGCAAGAAGACCAAGA 2407  
Qy 1334 AGGAGCCCTTCTGCTGCGCATCGAGCTGACCCCGCAAGTGAACCTGTGAGCCATCG 1393  
Db 2408 AGGAGCCCTTCTGCTGCGCATCGAGCTGACCCCGCAAGTGAACCTGTGAGCCATCG 2467  
Qy 1394 AGCTGCCGAGAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGA 1453  
Db 2468 AGCTGCCGAGAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGA 2527  
Qy 1454 ACTGGGCGAGCAGATCTTACCCCGGCATCAAGGTGCGCGAGCTGTGAAGCTGTGCGCG 1513  
Db 2528 ACTGGGCGAGCAGATCTTACCCCGGCATCAAGGTGCGCGAGCTGTGAAGCTGTGCGCG 2587  
Qy 1514 GCGCAAGGCCCTGACCGACATCGTGCCCCCTGAACGAGGAGCGAGCTGTGAGCTGGCG 1573  
Db 2588 GCGCAAGGCCCTGACCGACATCGTGCCCCCTGAACGAGGAGCGAGCTGTGAGCTGGCG 2647  
Qy 1574 AGAACCGGAGATCTGCGCGAGCCGTGACCGGCTGTACTACGACCCCGCAAGGACC 1633  
Db 2648 AGAACCGGAGATCTGCGCGAGCCGTGACCGGCTGTACTACGACCCCGCAAGGACC 2707  
Qy 1634 TGGTGGCGGAGATCCAGAAGCAGGCGCCACGACCTGAGTGACCTTACAGAGGAGC 1693  
Db 2708 TGGTGGCGGAGATCCAGAAGCAGGCGCCACGACCTGAGTGACCTTACAGAGGAGC 2767  
Qy 1694 CTTTCAAGAACCTGAGAACCGGCAAGTACGCAAGATGGCCACCGCCACCAACAGGAG 1753  
Db 2768 CTTTCAAGAACCTGAGAACCGGCAAGTACGCAAGATGGCCACCGCCACCAACAGGAG 2827  
Qy 1754 TGAACGACTGACCGAGGCGGTGCAAGAGATGCCATGGAGAGCATCGTGATCTGGGCA 1813  
Db 2828 TGAACGACTGACCGAGGCGGTGCAAGAGATGCCATGGAGAGCATCGTGATCTGGGCA 2887  
Qy 1814 AGACCCCAAGTTCGCTGCGCATCCAGAGGAGACCTGGGAGACCTGGTGAACCGACT 1873  
Db 2888 AGACCCCAAGTTCGCTGCGCATCCAGAGGAGACCTGGGAGACCTGGTGAACCGACT 2947  
Qy 1874 ACTGCGCGCACCTGGAATCCCGAGTGGGAGTTGTTGAACACCCCGCCCTGGTGAGC 1933  
Db 2948 ACTGCGCGCACCTGGAATCCCGAGTGGGAGTTGTTGAACACCCCGCCCTGGTGAGC 3007  
Qy 1934 TGTGTTACAGCTGGAGAGGAGGCCATCATCGGCGCGGAGACCTTCTACGTGGAGCGGCG 1993  
Db 3008 TGTGTTACAGCTGGAGAGGAGGCCATCATCGGCGCGGAGACCTTCTACGTGGAGCGGCG 3067  
Qy 1994 CGGCCAACCGGAGACCAAGATTCGCAAGGCGGCTACGTGACCGACCGGGCGCGGCA 2053  
Db 3068 CGGCCAACCGGAGACCAAGATTCGCAAGGCGGCTACGTGACCGACCGGGCGCGGCA 3127  
Qy 2054 AGATCGTGAAGCTGACCGAGACCAACACGAGAACCGAGCTGAGGCCATCCAGCTGG 2113  
Db 3128 AGATCGTGAAGCTGACCGAGACCAACACGAGAACCGAGCTGAGGCCATCCAGCTGG 3187  
Qy 2114 CCCTGCGAGACAGCGGCGAGCGAGTGAACATCGTGACCGACAGCGAGTACGCCCTGGGCA 2173  
Db 3188 CCCTGCGAGACAGCGGCGAGCGAGTGAACATCGTGACCGACAGCGAGTACGCCCTGGGCA 3247  
Qy 2174 TCATCCAGGCCCGCAGCCGCAAGAGCGAGAGCTGGTGAACACAGATCATCGAGCAGC 2233  
Db 3248 TCATCCAGGCCCGCAGCCGCAAGAGCGAGAGCTGGTGAACACAGATCATCGAGCAGC 3307  
Qy 2234 TGATCAAGAGGAGAGAGTGTACCTGAGCTGGGTGCGGCCCAACAGGGCATCGCGGCA 2293  
Db 3308 TGATCAAGAGGAGAGAGTGTACCTGAGCTGGGTGCGGCCCAACAGGGCATCGCGGCA 3367

```
Qy 2294 ACGAGCAGATCGACAAGCTGTGTAGCAGGAGCATCCCGCAAGGTGCTGTCTTGACGGCA 2353
Db 3368 ACGAGCAGATCGACAAGCTGTGTAGCAGGAGCATCCCGCAAGGTGCTGTCTTGACGGCA 3427
Qy 2354 TCGATGGCGGCATCGTGATCTTACCAAGTACATGGAGCACTTACGTGGGCAAGCGCGGCC 2413
Db 3428 TCGATGGCGGCATCGTGATCTTACCAAGTACATGGAGCACTTACGTGGGCAAGCGCGGCC 3487
Qy 2414 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451
Db 3488 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3525

RESULT 14
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBERG, Esmerita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 45
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_c
US-10-190-435-45

Query Match 97.8%; Score 2403.4; DB 15; Length 2457;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2439; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

Qy 7 GCCACCATGGCCGAGGCCATGAGCGCCACCGAGCGCCCAACATCTCTGATGAGCGAGC 66
Db 1 GCCACCATGGCCGAGGCCATGAGCGCCACCGAGCGCCCAACATCTCTGATGAGCGAGC 60
Qy 67 AACTTCAAGGGCCCGAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGGGCCAATC 126
Db 61 AACTTCAAGGGCCCGAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGGGCCAATC 120
Qy 127 GCCGCAACTGCGCGCCCGCCGCAAGAGGGCTGCTGGAGTGGCGGCAAGGGGCCAC 186
Db 121 GCCGCAACTGCGCGCCCGCCGCAAGAGGGCTGCTGGAGTGGCGGCAAGGGGCCAC 180
Qy 187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCCCTTCCC 246
Db 181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCCCTTCCC 240
Qy 247 CAGGGCAAGGCCCGCGAGTTCCCGAGGAGCAGAAACCGCGCAACAGCCCCACAGCGCG 306
Db 241 CAGGGCAAGGCCCGCGAGTTCCCGAGGAGCAGAAACCGCGCAACAGCCCCACAGCGCG 300
Qy 307 GAGCTGAGGTGCGCGGCAAAACCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db 301 GAGCTGAGGTGCGCGGCAAAACCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 360
Qy 367 CTGAACCTTCCCGCAGATCACCTGTGTGCGAGCGCCCTCTGTGTGAGCATCAAGGTGGCGGC 426
Db 361 CTGAACCTTCCCGCAGATCACCTGTGTGCGAGCGCCCTCTGTGTGAGCATCAAGGTGGCGGC 420
Qy 427 CAGATCAAGGAGGCCCTGTGTGAACCGGGCGCGAGCACCGTGTGTGGAGGAGATGAGC 486
Db 421 CAGATCAAGGAGGCCCTGTGTGAACCGGGCGCGAGCACCGTGTGTGGAGGAGATGAGC 480
```

```
Qy 487 CTGCCGCGCAAGTGAAGCCCAAGATGATCGCGGCATTCGGCGGCTTTCATCAAGGTGCGC 546
Db 481 CTGCCGCGCAAGTGAAGCCCAAGATGATCGCGGCATTCGGCGGCTTTCATCAAGGTGCGC 540
Qy 547 CAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCACTCGGCACCGTGTGATC 606
Db 541 CAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCACTCGGCACCGTGTGATC 600
Qy 607 GGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCGTG 666
Db 601 GGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCGTG 660
Qy 667 AACTTCCCATTCAGCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGCATCGACGGC 726
Db 661 AACTTCCCATTCAGCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGCATCGACGGC 720
Qy 727 CCAAGGTGAAGCAGTGGCCCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGC 786
Db 721 CCAAGGTGAAGCAGTGGCCCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGC 780
Qy 787 GAGGAGTGGAGAGGAGGCAAGATCACCAAGATCGGCCCGAGAAACCCCTCAACACC 846
Db 781 GAGGAGTGGAGAGGAGGCAAGATCACCAAGATCGGCCCGAGAAACCCCTCAACACC 840
Qy 847 CCGGTGTTCCGCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db 841 CCGGTGTTCCGCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy 907 GAGCTGAACCAAGCGACCCAGGACTTCTGGAGAGTGCAGTGGGCATCCCCACCCCGCC 966
Db 901 GAGCTGAACCAAGCGACCCAGGACTTCTGGAGAGTGCAGTGGGCATCCCCACCCCGCC 960
Qy 967 GGCTGAAGAGAGAGAGGCTGACCGTGTGACGTGGCGAGCGCTTACTTACGGTG 1026
Db 961 GGCTGAAGAGAGAGAGGCTGACCGTGTGACGTGGCGAGCGCTTACTTACGGTG 1020
Qy 1027 CCGCTGACGAGGACTTCCGCAAGTACACCGCTTACACCATCCCGAGCATCAACAAACAG 1086
Db 1021 CCGCTGACGAGGACTTCCGCAAGTACACCGCTTACACCATCCCGAGCATCAACAAACAG 1080
Qy 1087 ACCCGCGCATTCGCTACCAAGTACAACTGTGCTCCCGAGGCTGGAGGGCAGCCCCAGC 1146
Db 1081 ACCCGCGCATTCGCTACCAAGTACAACTGTGCTCCCGAGGCTGGAGGGCAGCCCCAGC 1140
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAACCCCGAGTC 1206
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAACCCCGAGTC 1200
Qy 1207 GTGATCTACCA-----GGCCCCCTGTAGCTGGGCGAGCGACCTGGAGATCGGCGCAGC 1260
Db 1201 GTGATCTACCAAGTACATGGAGCACTGTAGCTGGGCGAGCGACCTGGAGATCGGCGCAGC 1260
Qy 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCAGCTGTGCTGGCTGGGGCTTTCACACCCCGAC 1320
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCAGCAGCTGTGCTGGCTGGGGCTTTCACACCCCGAC 1320
Qy 1321 AAGAGCACCAGAGAGGAGCCCTTCTCTGCCAT-----CGAGCTGCACCCCGACAG 1374
Db 1321 AAGAGCACCAGAGAGGAGCCCTTCTCTGGTGGGGCTTACGAGCTGCACCCCGACAG 1380
Qy 1375 TGGACCGTGCAGCCCTCGAGCTGCGCGAGAGAGAGCTGGACCGGTGAACGACATCCAG 1434
Db 1381 TGGACCGTGCAGCCCTCGAGCTGCGCGAGAGAGAGCTGGACCGGTGAACGACATCCAG 1440
Qy 1435 AAGTGTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTTACCCCGCATCAAGGTGCGCCAG 1494
Db 1441 AAGTGTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTTACCCCGCATCAAGGTGCGCCAG 1500
Qy 1495 CTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCTGTGCGCCCTGACCGAGGAG 1554
Db 1501 CTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCTGTGCGCCCTGACCGAGGAG 1560
```

```
QY 1555 GCCAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTCGCGAGCCCGCTGCAACGCGCTGTAC 1614
Db 1561 GCCAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGGCGAGCCCGTGTCAACGCGCTGTAC 1620
QY 1615 TAGGACCCAGCAAGGACTGTGTGGCGGAGATCCAGAGAGAGGCGCAAGACGAGTGGACC 1674
Db 1621 TAGGACCCAGCAAGGACTGTGTGGCGGAGATCCAGAGAGAGGCGCGAGACCGAGTGGACC 1680
QY 1675 TACGAGATCTACAGGAGCCCTTCAAGAACCTTGAGACCGGCAAGTACGCGCAAGATGGC 1734
Db 1681 TACGAGATCTACAGGAGCCCTTCAAGAACCTTGAGAACCGGCAAGTACGCGCAAGATGGC 1740
QY 1735 ACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAGAGATCCGCATGGAG 1794
Db 1741 ACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAGATCCGCATGGAG 1800
QY 1795 AGCATGTGTATCTGGGGCAAGACCCCAAGTTTCGCTGCCCATCCAGAGAGACCTGG 1854
Db 1801 AGCATGTGTATCTGGGGCAAGACCCCAAGTTTCGCTGCCCATCCAGAGAGACCTGG 1860
QY 1855 GAGACCTGTGTACCGACTACTCTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1914
Db 1861 GAGACCTGTGTACCGACTACTCTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1920
QY 1915 ACCCCCCCTGCTGAAGCTGTGTATACAGCTGGAGAGGAGCCCATCATCGCGCCGAG 1974
Db 1921 ACCCCCCCTGCTGAAGCTGTGTATACAGCTGGAGAGGAGCCCATCATCGCGCCGAG 1980
QY 1975 ACCTTTCTAGTGGACGGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2034
Db 1981 ACCTTTCTAGTGGACGGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2040
QY 2035 ACCGACGGGGCGGCGAGAGATCTGAGCTGACCGAGACCGACCAACGAGAGACCGAG 2094
Db 2041 ACCGACGGGGCGGCGAGAGATCTGAGCTGACCGAGACCGACCAACGAGAGACCGAG 2100
QY 2095 CTGCAAGGCCCATCAGCTGGCCCTGAGGACAGCGGAGCGAGGTGAACATCTGTGACCGAC 2154
Db 2101 CTGCAAGGCCCATCAGCTGGCCCTGAGGACAGCGGAGCGAGGTGAACATCTGTGACCGAC 2160
QY 2155 AGCCAGTACGCCCTGGGCATCATCAGGCCAGCCGACCAAGAGAGAGCGAGCTGTGTG 2214
Db 2161 AGCCAGTACGCCCTGGGCATCATCAGGCCAGCCGACCAAGAGAGAGCGAGCTGTGTG 2220
QY 2215 ACCAGATCATCAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTGGCCGCC 2274
Db 2221 ACCAGATCATCAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTGGCCGCC 2280
QY 2275 CACAAGGGCATCGCGGCAACGAGCAGATCGACAAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
Db 2281 CACAAGGGCATCGCGGCAACGAGCAGATCGACAAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
QY 2335 GTGCTGTTCTGGACGGCATCATCGCGGCATCGTGTATCCAGTATACGAGACACCTG 2394
Db 2341 GTGCTGTTCTGGACGGCATCATCGCGGCATCGTGTATCCAGTATACGAGACACCTG 2400
QY 2395 TACGTGGGAGCGGGCGCCCTAGGATCGATTTAAAGCTTCCCGGGCTAGCACCGGT 2451
Db 2401 TACGTGGGAGCGGGCGCCCTAGGATCGATTTAAAGCTTCCCGGGCTAGCACCGGT 2457
```

```
RESULT 15
US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
```

```
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

Query Match 97.8%; Score 2403.4; DB 16; Length 2457;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2439; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

QY 7 GCCACCATGGCGAGGCCATGAGCCAGCGCCACCGAGCGCAACATCTGTGCGAGCGCAGC 66
Db 1 GCCACCATGGCGAGGCCATGAGCCAGCGCCACCGAGCGCAACATCTGTGCGAGCGCAGC 60
QY 67 AACTTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATC 126
Db 61 AACTTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATC 120
QY 127 GCCGCAACTGCGCGCCCCCGCAAGAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCAC 186
Db 121 GCCGCAACTGCGCGCCCCCGCAAGAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCAC 180
QY 187 CAGATGAAGAGTGTGACCGAGCGCCAGGCCAACTTCTTCGCGGAGGACCTGGCCCTTCCC 246
Db 181 CAGATGAAGAGTGTGACCGAGCGCCAGGCCAACTTCTTCGCGGAGGACCTGGCCCTTCCC 240
QY 247 CAGGCAAGGGCCCGCGAGTTCGCCAGCGAGCAGAACCGCGCAACAGCCGCCACAGAGCGC 306
Db 241 CAGGCAAGGGCCCGCGAGTTCGCCAGCGAGCAGAACCGCGCAACAGCCGCCACAGAGCGC 300
QY 307 GAGCTGCAAGTGTGCGGGCGCAACCCCGCAGCGAGCGCGCGCGCGCGAGCGCCAGGCGACC 366
Db 301 GAGCTGCAAGTGTGCGGGCGCAACCCCGCAGCGAGCGCGCGCGCGCGAGCGCCAGGCGACC 360
QY 367 CTGAACCTTCCCGCAGATCACCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTGGCGGC 426
Db 361 CTGAACCTTCCCGCAGATCACCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTGGCGGC 420
QY 427 CAGATCAAGGAGGCCCTCTGTGACACCGCGCGCGAGCACACCGTCTCGAGAGAGATGAGC 486
Db 421 CAGATCAAGGAGGCCCTCTGTGACACCGCGCGCGAGCACACCGTCTCGAGAGAGATGAGC 480
QY 487 CTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAAGGTGGCGC 546
Db 481 CTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAAGGTGGCGC 540
QY 547 CAGTACGACCGAGATCTGTATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTGTGATC 606
Db 541 CAGTACGACCGAGATCTGTATCGAGATCTGGCGCAAGAGGCCATCGGCACCGTGTGATC 600
QY 607 GGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG 666
Db 601 GGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG 660
QY 667 AACTTTCCCATCAGCGCCCATCGAGACCGTGGCCGCTGAAGCTGAAGCCCGCGCATGACCGC 726
Db 661 AACTTTCCCATCAGCGCCCATCGAGACCGTGGCCGCTGAAGCTGAAGCCCGCGCATGACCGC 720
QY 727 CCCAAGGTGAAGAGTGGCGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCG 786
Db 721 CCCAAGGTGAAGAGTGGCGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCG 780
QY 787 GAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTTACAAACACC 846
Db 781 GAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTTACAAACACC 840
QY 847 CCCGTGTTCCGCATCAAGAGAGAGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db 847 CCCGTGTTCCGCATCAAGAGAGAGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
```

Db 841 CCCGTGTTCCCATCAAGAGAGGACAGCACCAGTGGCGGCAAGCTGGTGGACTTCCGC 900  
Qy 907 GAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGACGCTGGGCATCCCCACCCCGCC 966  
Db 901 GAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGACGCTGGGCATCCCCACCCCGCC 960  
Qy 967 GGCCTGAAGAGAGAGAGGCTGACCGTGTGAGAGCTGGGGGAGCGCTACTTACAGCGTG 1026  
Db 961 GGCCTGAAGAGAGAGAGGCTGACCGTGTGAGAGCTGGGGGAGCGCTACTTACAGCGTG 1020  
Qy 1027 CCCCTGGACGAGGACTTCCGCAAGTACACCCGCTTCCACATCCCGAGCATCAACAACGAG 1086  
Db 1021 CCCCTGGACGAGGACTTCCGCAAGTACACCCGCTTCCACATCCCGAGCATCAACAACGAG 1080  
Qy 1087 ACCCCCGGCATCCGCTACAGTACAAAGTGTGCTGCCCAAGGGCTGGAAGGGCAGCCCGACG 1146  
Db 1081 ACCCCCGGCATCCGCTACAGTACAAAGTGTGCTGCCCAAGGGCTGGAAGGGCAGCCCGACG 1140  
Qy 1147 ATCTTCCAGAGGAGATGACCAAGATCTGGAGCGCTTCCGCGCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCCAGAGGAGATGACCAAGATCTGGAGCGCTTCCGCGCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACCA- ----GGCCCCCTGTACTGTGGCGAGCGACTGGAGATCGGCCAGC 1260  
Db 1201 GTGATCTACCAAGTACGACCTGTGACGTGGCGAGCGACTGGAGATCGGCCAGC 1260  
Qy 1261 CGCGCCAAGATCGAGGAGCTCGCAAGCACTGTGCTGGGCTTCCACCAACCCCGAC 1320  
Db 1261 CGCGCCAAGATCGAGGAGCTCGCAAGCACTGTGCTGGGCTTCCACCAACCCCGAC 1320  
Qy 1321 AAGAAGCAACAGAGGAGCCCTTCTGCTGCCAT- ----CGAGTGCACCCCGCAAG 1374  
Db 1321 AAGAAGCAACAGAGGAGCCCTTCTGCTGGATGGGCTACGAGTGCACCCCGCAAG 1380  
Qy 1375 TGGACCTGACGAGCCATCGAGCTCCGAGAGGAGCTGACCGTGAACGACATCCAG 1434  
Db 1381 TGGACCTGACGAGCCATCGAGCTCCGAGAGGAGCTGACCGTGAACGACATCCAG 1440  
Qy 1435 AAGCTGTGGGCAAGCTGAATGGGGCCAGCAGATCTACCCCGGCATCAAGTGGCCAG 1494  
Db 1441 AAGCTGTGGGCAAGCTGAATGGGGCCAGCAGATCTACCCCGGCATCAAGTGGCCAG 1500  
Qy 1495 CTGTGCAAGCTGTGCGCGGCGCAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAG 1554  
Db 1501 CTGTGCAAGCTGTGCGCGGCGCAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAG 1560  
Qy 1555 GCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGCGGAGCCCGTGACGCGGTGTAC 1614  
Db 1561 GCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGCGGAGCCCGTGACGCGGTGTAC 1620  
Qy 1615 TACGACCCAGCAAGGAGCTGTGGCGGAGATCCAGAGCAGGGCCAGCAGTGGACC 1674  
Db 1621 TACGACCCAGCAAGGAGCTGTGGCGGAGATCCAGAGCAGGGCCAGCAGTGGACC 1680  
Qy 1675 TACGAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1734  
Db 1681 TACGAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740  
Qy 1735 ACCGCCACACCAAGAGGAGTGAAGCTGACCGAGGCGCTGCAAGAGATCGCCATGGAG 1794  
Db 1741 ACCGCCACACCAAGAGGAGTGAAGCTGACCGAGGCGCTGCAAGAGATCGCCATGGAG 1800  
Qy 1795 AGCATGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTCCCATCCAGAGGAGACCTGG 1854  
Db 1801 AGCATGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTCCCATCCAGAGGAGACCTGG 1860  
Qy 1855 GAGACCTGTGGACCGACTACTTGGCAGGCCCACTTGGATCCCGAGTGGAGTTCGTGAAC 1914  
Db 1861 GAGACCTGTGGACCGACTACTTGGCAGGCCCACTTGGATCCCGAGTGGAGTTCGTGAAC 1920  
Qy 1915 ACCCCCCCTGTGTGAAGCTGTGTATCCAGCTGAGAGGAGCCCATCATCGGCGCGGAG 1974  
Db 1921 ACCCCCCCTGTGTGAAGCTGTGTATCCAGCTGAGAGGAGCCCATCATCGGCGCGGAG 1980

Qy 1975 ACCTTTTACGTGGACGGCGCCCAACCGCGAGACCAGATCGGCAAGCGCGGTACGTG 2034  
Db 1981 ACCTTTTACGTGGACGGCGCCCAACCGCGAGACCAGATCGGCAAGCGCGGTACGTG 2040  
Qy 2035 ACCGACCGGGCCCGGCAAGAGATCGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2094  
Db 2041 ACCGACCGGGCCCGGCAAGAGATCGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2100  
Qy 2095 CTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGAGGTGAACATCGTGACCGAC 2154  
Db 2101 CTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGAGGTGAACATCGTGACCGAC 2160  
Qy 2155 AGCCAGTACGCCCTTGGGCATCATCCAGGCCAGCCCGACAAAGAGCGAGAGCGAGTGGTG 2214  
Db 2161 AGCCAGTACGCCCTTGGGCATCATCCAGGCCAGCCCGACAAAGAGCGAGAGCGAGTGGTG 2220  
Qy 2215 AACGAGATCATCGAGGAGCTGTATCAAGAGAGAGAGGTGTACCTGAGCTGGGTGCCGCC 2274  
Db 2221 AACGAGATCATCGAGGAGCTGTATCAAGAGAGAGAGGTGTACCTGAGCTGGGTGCCGCC 2280  
Qy 2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334  
Db 2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340  
Qy 2335 GTGCTGTTCTTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGAGAGACCTG 2394  
Db 2341 GTGCTGTTCTTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGAGAGACCTG 2400  
Qy 2395 TACGTGGGCGAGCGCGGCCCTTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 2451  
Db 2401 TACGTGGGCGAGCGCGGCCCTTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 2457

Search completed: June 2, 2005, 04:23:48  
Job time : 1415.9 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2005, 23:20:00 ; Search time 1267.33 Seconds  
(without alignments)  
11504.726 Million cell updates/sec

Title: US-09-610-313B-31  
Perfect score: 2463  
Sequence: 1 gtcagccaccatgcccga.....gggctagaccggtgaattc 2463

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2463	100.0	2463	6 ABL39960	AbL39960 Synthetic
2	2463	100.0	2463	12 ADM73765	Adm73765 HIV-1 pol
3	2455.4	99.7	2457	8 ACA03547	Aca03547 Synthetic
4	2455.4	99.7	2457	10 ADC13265	Adc13265 DNA of HI
5	2442.2	99.2	2469	6 ABL39959	AbL39959 Synthetic
6	2442.2	99.2	2469	12 ADM73764	Adm73764 HIV-1 pol
7	2436.2	98.9	2457	6 ABL39961	AbL39961 Synthetic
8	2436.2	98.9	2457	12 ADM73766	Adm73766 HIV-1 pol
9	2430.2	98.7	2457	8 ACA03548	Aca03548 Synthetic
10	2430.2	98.7	2457	10 ADC13266	Adc13266 DNA of HI
11	2422.6	98.4	2445	8 ACA03546	Aca03546 Synthetic
12	2422.6	98.4	2445	10 ADC13284	Adc13284 DNA of HI
13	2415.6	98.1	3930	10 ADC13230	Adc13230 DNA of HI
14	2414	98.0	3930	10 ADC13231	Adc13231 DNA of HI
15	2414	98.0	3930	10 ADC13232	Adc13232 DNA of HI
16	2414	98.0	5184	8 ACA03591	Aca03591 Synthetic
17	2414	98.0	5184	10 ADC13279	Adc13279 DNA of HI
18	2383.6	96.8	3531	10 ADC13234	Adc13234 DNA of HI
19	2382	96.7	3537	10 ADC13236	Adc13236 DNA of HI
20	2381	96.7	5145	8 ACA03521	Aca03521 Synthetic

# ALIGNMENTS

## RESULT 1

ABL39960

ID ABL39960 standard; DNA; 2463 BP.

XX ABL39960;

XX 15-MAY-2002 (first entry)

DE Synthetic construct PR975YM SEQ ID NO:31.

XX Human immunodeficiency virus type C; antigenic HIV type C protein;  
immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;  
immunostimulant; gene therapy; gene; db.

OS Human immunodeficiency virus; type C.  
XX Synthetic.

PN WO200204493-A2.

XX 17-JAN-2002.

XX 05-JUL-2001; 2001WO-US021241.

PR 05-JUL-2000; 2000US-00610313.

XX (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2002-154920/20.

XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.

XX Claim 1; Fig 9; 233pp; English.

XX The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding gag, pol, vif, vpr, tat, rev, vpu, env or nef (i). (i) have immunostimulant activity and can be used in gene therapy.

21 2381 96.7 5145 10 ADC13233 Adc13233 DNA of HI  
22 2371 96.3 3538 10 ADC13235 Adc13235 DNA of HI  
23 2322.6 94.3 3607 8 ACA03551 Aca03551 Synthetic  
24 2322.6 94.3 3607 10 ADC13269 Adc13269 DNA of HI  
25 2322.6 94.3 3624 8 ACA03550 Aca03550 Synthetic  
26 2322.6 94.3 3624 10 ADC13268 Adc13268 DNA of HI  
27 2304.4 93.6 3597 8 ACA03549 Aca03549 Synthetic  
28 2304.4 93.6 3597 10 ADC13267 Adc13267 DNA of HI  
29 2142 87.0 2466 8 ACA03542 Aca03542 Synthetic  
30 2142 87.0 2466 8 ACC78506 Acc78506 HIV p2Pol  
31 2122.8 86.2 2472 8 ACA03543 Aca03543 Synthetic  
32 2122.8 86.2 2472 8 ACC78507 Acc78507 HIV p2Pol  
33 2115.2 85.9 2460 8 ACA03541 Aca03541 Synthetic  
34 2115.2 85.9 2460 8 ACC78505 Acc78505 HIV p2Pol  
35 2114.2 85.8 3564 8 ACC78488 Acc78488 HIV GagPo  
36 2114.2 85.8 3564 8 ACC78489 Acc78489 HIV GagPo  
37 2113.6 85.8 4716 8 ACA03522 Aca03522 Synthetic  
38 2113.6 85.8 4716 10 ADC13238 Adc13238 DNA of HI  
39 2109.8 85.7 3999 8 ACC78484 Acc78484 HIV GagCo  
40 2108.2 85.6 3999 8 ACC78485 Acc78485 HIV GagCo  
41 2108.2 85.6 3999 8 ACC78486 Acc78486 HIV GagCo  
42 2108.2 85.6 5283 8 ACA03584 Aca03584 Synthetic  
43 2108.2 85.6 5283 8 ACC78529 Acc78529 HIV TatRe  
44 2108 85.6 4713 8 ACA03592 Aca03592 Synthetic  
45 2108 85.6 4713 10 ADC13280 Adc13280 DNA of HI







```
QY 781 ATCTGCGAGAGATGGAGAGAGGGGCAAGATCACCAAGATCGSCCCCGAGAGACCCCTAC 840
DB 781 ATCTGCGAGAGATGGAGAGAGGGGCAAGATCACCAAGATCGSCCCCGAGAGACCCCTAC 840
QY 841 AACACCCCCCGTGTTCGCCATCAAGAAGAGAGAGAGACCAAGTGGCGCAAGCTGGTGGAC 900
DB 841 AACACCCCCCGTGTTCGCCATCAAGAAGAGAGAGAGACCAAGTGGCGCAAGCTGGTGGAC 900
QY 901 TTCCGCGAGCTGAACAAGCGGACCCAGGAATTCTTGGAGGTGAGCTGGGCATCCCCAC 960
DB 901 TTCCGCGAGCTGAACAAGCGGACCCAGGAATTCTTGGAGGTGAGCTGGGCATCCCCAC 960
QY 961 CCCGCGGCTGAAGAGAGAGAGAGGTGACCGTGTGACGTGGCGAGCGCTACTTC 1020
DB 961 CCCGCGGCTGAAGAGAGAGAGAGGTGACCGTGTGACGTGGCGAGCGCTACTTC 1020
QY 1021 AGCGTGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGCATCAAC 1080
DB 1021 AGCGTGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGCATCAAC 1080
QY 1081 AACGAGACCCCGGCTATCCGCTACAGTA CAACGTGTCTGCCAGGGCTGGAGGGGAGC 1140
DB 1081 AACGAGACCCCGGCTATCCGCTACAGTA CAACGTGTCTGCCAGGGCTGGAGGGGAGC 1140
QY 1141 CCCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCCCGCAACCCC 1200
DB 1141 CCCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCCCGCAACCCC 1200
QY 1201 GAGATCGTGATCTACCAAGGCCCCCTGTACGTGGGAGACCTGGAGATCGGCCAGCAC 1260
DB 1201 GAGATCGTGATCTACCAAGGCCCCCTGTACGTGGGAGACCTGGAGATCGGCCAGCAC 1260
QY 1261 CGCGCCAAGATCGAGGAGCTGGGCAAGCACTGTGCTGGCTGGGGCTTCAACAACCCCGAC 1320
DB 1261 CGCGCCAAGATCGAGGAGCTGGGCAAGCACTGTGCTGGCTGGGGCTTCAACAACCCCGAC 1320
QY 1321 AAGAGCACCAGAGGAGGCCCCCTTCTGTGATGGCTACGAGCTGCAACCCCGAAG 1380
DB 1321 AAGAGCACCAGAGGAGGCCCCCTTCTGTGATGGCTACGAGCTGCAACCCCGAAG 1380
QY 1381 TGGACCGTGCAGGCCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440
DB 1381 TGGACCGTGCAGGCCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440
QY 1441 AAGCTGTGGGCAAGCTGAATCTGGGCGAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
DB 1441 AAGCTGTGGGCAAGCTGAATCTGGGCGAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
QY 1501 CTGTGAAGCTGTGCGCGGCGCAAGGCCCTGACCGACATCTGTCGCCCTGACCGGAG 1560
DB 1501 CTGTGAAGCTGTGCGCGGCGCAAGGCCCTGACCGACATCTGTCGCCCTGACCGGAG 1560
QY 1561 GCGAGCTGAGCTGGCGGAGAACCGCGAGATCTGCGGAGCCCGTGCACGGCGTGTAC 1620
DB 1561 GCGAGCTGAGCTGGCGGAGAACCGCGAGATCTGCGGAGCCCGTGCACGGCGTGTAC 1620
QY 1621 TAGAACCCCAAGAGGACTGTGGTGGCGGAGATCCAGAAGCAGGGGCGACGACCGATGACC 1680
DB 1621 TAGAACCCCAAGAGGACTGTGGTGGCGGAGATCCAGAAGCAGGGGCGACGACCGATGACC 1680
QY 1681 TACAGATCTACGAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATCGCG 1740
DB 1681 TACAGATCTACGAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATCGCG 1740
QY 1741 ACCGCCACCAACGACGCTGAAGAGCTGACCGAGGCCGTGCAAGAGATCGCCATGGAG 1800
DB 1741 ACCGCCACCAACGACGCTGAAGAGCTGACCGAGGCCGTGCAAGAGATCGCCATGGAG 1800
QY 1801 AGATCTGTATCTGGGGCAAGACCCCAAGTTCGGCTGCGCATCCGAAGAGGACCTGG 1860
DB 1801 AGATCTGTATCTGGGGCAAGACCCCAAGTTCGGCTGCGCATCCGAAGAGGACCTGG 1860
```

```
QY 1861 GAGACCTGTGTGAGCCGACTACTGCGAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1920
DB 1861 GAGACCTGTGTGAGCCGACTACTGCGAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1920
QY 1921 ACCCCGCCCTGTGTGAAGCTGTGTACAGCTGGAGAGAGGCCATCATCGGCGCCGAG 1980
DB 1921 ACCCCGCCCTGTGTGAAGCTGTGTACAGCTGGAGAGAGGCCATCATCGGCGCCGAG 1980
QY 1981 ACCTTCTAGTGAAGCGCGCCCAACCGCGAGACCAAGATCGCAAGGCCGGCTACGTG 2040
DB 1981 ACCTTCTAGTGAAGCGCGCCCAACCGCGAGACCAAGATCGCAAGGCCGGCTACGTG 2040
QY 2041 ACCGACCGGCGCGCGAGAGATCTGAGCTTCAACCGAGACCAACACCAAGAGACCGAG 2100
DB 2041 ACCGACCGGCGCGCGAGAGATCTGAGCTTCAACCGAGACCAACACCAAGAGACCGAG 2100
QY 2101 CTGACGGCCATTCAGCTGGCCCTGCGAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCAG 2160
DB 2101 CTGACGGCCATTCAGCTGGCCCTGCGAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCAG 2160
QY 2161 AGCAGTACGCGCTGGGCTATCCAGGCCCGCCAGCAAGAGCGAGAGCTGGT 2220
DB 2161 AGCAGTACGCGCTGGGCTATCCAGGCCCGCCAGCAAGAGCGAGAGCTGGT 2220
QY 2221 AACAGATCATTCAGAGAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCCCGCC 2280
DB 2221 AACAGATCATTCAGAGAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCCCGCC 2280
QY 2281 CACAAGGCCATCGCGGCGCAACGAGAGATCGACAAGCTGGTGAAGGGGATCCGCAAG 2340
DB 2281 CACAAGGCCATCGCGGCGCAACGAGAGATCGACAAGCTGGTGAAGGGGATCCGCAAG 2340
QY 2341 GTCTGTTCTGACCGGCTATCGATGGGCGATCTACCATCTACCATGAGAGAGCTG 2400
DB 2341 GTCTGTTCTGACCGGCTATCGATGGGCGATCTACCATCTACCATGAGAGAGCTG 2400
QY 2401 TAGTGGGCGAGCGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
DB 2401 TAGTGGGCGAGCGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
QY 2461 TTC 2463
DB 2461 TTC 2463

RESULT 3
ACA03547
ID ACA03547 standard; DNA; 2457 BP.
XX
AC ACA03547;
XX
AC ACA03547;
XX
DT 22-MAY-2003 (first entry)
XX
XX Synthetic DNA encoding immunogenic HIV peptide #30.
DE Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
XX Synthetic.
XX
XX WO2003004657-A1.
XX
XX 16-JAN-2003.
XX
XX 05-JUL-2002; 2002WO-US021421.
XX
XX 05-JUL-2001; 2001US-0303192P.
XX
XX 31-AUG-2001; 2001US-0316860P.
XX
XX 16-JAN-2002; 2002US-0349728P.
XX
XX 16-JAN-2002; 2002US-0349793P.
XX
XX 16-JAN-2002; 2002US-0349871P.
XX
```



```
Db 1681 TACCAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCGCAAGATCGCG 1740
Qy 1741 ACCGCCACACCAACGACGCTGACGACGCTGACGAGGCGCTGACGAGATCGCCATGGAG 1800
Db 1741 ACCGCCACACCAACGACGCTGACGACGCTGACGAGGCGCTGACGAGATCGCCATGGAG 1800
Qy 1801 AGCATCGTGTCTGGGGCAAGACCCCAAGTTCGCGCTGCCCTCCATCCAGAGGAGACCTGG 1860
Db 1801 AGCATCGTGTCTGGGGCAAGACCCCAAGTTCGCGCTGCCCTCCATCCAGAGGAGACCTGG 1860
Qy 1861 GAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAAC 1920
Db 1861 GAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAAC 1920
Qy 1921 ACCCCCCCTGTGTGAGCTGTGTACACGCTGGAGAGGCCCATCATCGCGCGCGAG 1980
Db 1921 ACCCCCCCTGTGTGAGCTGTGTACACGCTGGAGAGGCCCATCATCGCGCGCGAG 1980
Qy 1981 ACCTTCTACGTGGACGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGTACGTTG 2040
Db 1981 ACCTTCTACGTGGACGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGTACGTTG 2040
Qy 2041 ACCGACCGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Db 2041 ACCGACCGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Qy 2101 CTGACGCGCATTCAGCTGGCCCTGCAGACAGCGGACGAGGTGAACATCGTGACCGAC 2160
Db 2101 CTGACGCGCATTCAGCTGGCCCTGCAGACAGCGGACGAGGTGAACATCGTGACCGAC 2160
Qy 2161 AGCCAGTACGCCCTGGGCGCATCATCCAGGCCAGCCGACAGAGCGAGAGCGTGGTG 2220
Db 2161 AGCCAGTACGCCCTGGGCGCATCATCCAGGCCAGCCGACAGAGCGAGAGCGTGGTG 2220
Qy 2221 AACCAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGCCCGCC 2280
Db 2221 AACCAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGCCCGCC 2280
Qy 2281 CACNAGGGCATCGGCGCAAGCAGACATCGACNAGCTGTGTGACNAGGGCATCCGCAAG 2340
Db 2281 CACNAGGGCATCGGCGCAAGCAGACATCGACNAGCTGTGTGACNAGGGCATCCGCAAG 2340
Qy 2341 GTGCTGTCTCTGGACGCGCTGATGGCGGATCGTGATCTACCAATGATGAGACGCTG 2400
Db 2341 GTGCTGTCTCTGGACGCGCTGATGGCGGATCGTGATCTACCAATGATGAGACGCTG 2400
Qy 2401 TAGTGGGCAAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457
Db 2401 TAGTGGGCAAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457
```

RESULT 4  
ADCI3265  
ID ADCI3265 standard; DNA; 2457 BP.

AC ACDC13265;

XX 18-DEC-2003 (first entry)

XX DNA of HIV construct p2Pol-opt-YM\_C SEQ ID NO 44.

XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX Human immunodeficiency virus.

XX WO2003004620-A2.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.

```
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
XX (UYST-) UNIV STELLENBOSCH.  
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX WPI; 2003-221593/21.  
XX  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX  
XX Disclosure; Fig 41; 301pp; English.  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramuscularly, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX  
XX Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;
```

```
Query Match 99.7%; Score 2455.4; DB 10; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 3.5e-296;  
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GTCAGCGCCACCATGCGCGAGGCCATGAGCGCCACCGAGCGCAACATCTCTGATGCGAG 60  
Db 1 GTCAGCGCCACCATGCGCGAGGCCATGAGCGCCACCGAGCGCAACATCTCTGATGCGAG 60  
Qy 61 CGCAGCAACTTCAAGGGGCCCAAGCGCATCATCAAGTGTCTCACTCGCGCAAGAGGGGC 120  
Db 61 CGCAGCAACTTCAAGGGGCCCAAGCGCATCATCAAGTGTCTCACTCGCGCAAGAGGGGC 120  
Qy 121 CACATCGCCCGCAACTGCGCGCGCCCGCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCCGCAACTGCGCGCGCCCGCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Qy 181 GGCACACAGATGAAGGACTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 181 GGCACACAGATGAAGGACTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Qy 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCGCAGCGAGCAGAACCGCGCAACAGCCCCCACC 300  
Db 241 TTCCCCCAGGGCAAGGCCCGCGAGTTCCCGCAGCGAGCAGAACCGCGCAACAGCCCCCACC 300  
Qy 301 AGCGCGGAGCTGCAGGTGCGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCG 360  
Db 301 AGCGCGGAGCTGCAGGTGCGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCG 360  
Qy 361 GGCAACCTGAATTCCTCCCGCAGATCAACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCG 420  
Db 361 GGCAACCTGAATTCCTCCCGCAGATCAACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCG 420  
Qy 421 GCGCGCGCAGATCAAGGAGGCGCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 421 GCGCGCGCAGATCAAGGAGGCGCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Qy 481 ATGAGCCTGCGCGCAAGTGAAGGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Db 481 ATGAGCCTGCGCGCAAGTGAAGGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
Qy 541 GTGCGCGCAGTACGACCGAGATCTGATGATCGAGATCTGCGGCAAGAGGCGCGCGCGCGCGCG 600  
Db 541 GTGCGCGCAGTACGACCGAGATCTGATGATCGAGATCTGCGGCAAGAGGCGCGCGCGCGCGCG 600
```

```
Db 541 GTGCGCCAGTACGACCATCTCGATCGAGATCTCGGCGAAGAGGCCATCGGCACCGTG 600
Qy 601 CTGATCGGCCACCCCGTGAACATCATTCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCCACCCCGTGAACATCATTCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAATCTCCCATCAGCCCATCAGACCGTGCCTGGAAGCTGAAGCCCGGCATG 720
Db 661 ACCCTGAATCTCCCATCAGCCCATCAGACCGTGCCTGGAAGCTGAAGCCCGGCATG 720
Qy 721 GACGCCCCAAGGTGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCC 780
Db 721 GACGCCCCAAGGTGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCC 780
Qy 781 ATCTCGAGGAGATGAGAAAGAGGGCAAGATACCAAGATCGGCCCGGAGAACCCCTAC 840
Db 781 ATCTCGAGGAGATGAGAAAGAGGGCAAGATACCAAGATCGGCCCGGAGAACCCCTAC 840
Qy 841 AACACCCCGGTTCGCCCATCAAGAGAGAGACACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCGGTTCGCCCATCAAGAGAGAGACACCAAGTGGCGCAAGCTGGTGGAC 900
Qy 901 TTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCAC 960
Db 901 TTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCAC 960
Qy 961 CCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGGACGCTGGGCGACGCTACTTC 1020
Db 961 CCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGGACGCTGGGCGACGCTACTTC 1020
Qy 1021 AGCGTGCCTTGGAGGAGACTTCGGGAAGTACACCGCTTACCATCCCGAGCATCAAC 1080
Db 1021 AGCGTGCCTTGGAGGAGACTTCGGGAAGTACACCGCTTACCATCCCGAGCATCAAC 1080
Qy 1081 AACGAGACCCCGGCTACCGTACCAAGTACCAAGTGTCCCGCGGCTGGAAGGGGAGC 1140
Db 1081 AACGAGACCCCGGCTACCGTACCAAGTACCAAGTGTCCCGCGGCTGGAAGGGGAGC 1140
Qy 1141 CCGAGCATCTTCCAGAGCAGATCAACAGATCTCTGGAGCCCTTCGCGCGCGCAACCCC 1200
Db 1141 CCGAGCATCTTCCAGAGCAGATCAACAGATCTCTGGAGCCCTTCGCGCGCGCAACCCC 1200
Qy 1201 GAGATCGTGATCTACGAGCCCTCTGATCGTGGGACGACCTTGAGATCGGCGAGCAC 1260
Db 1201 GAGATCGTGATCTACGAGCCCTCTGATCGTGGGACGACCTTGAGATCGGCGAGCAC 1260
Qy 1261 CGCCCAAGATCGAGGAGCTGGCAAGCACCTGCTGGCTGGGCTTCAACACCCCGAC 1320
Db 1261 CGCCCAAGATCGAGGAGCTGGCAAGCACCTGCTGGCTGGGCTTCAACACCCCGAC 1320
Qy 1321 AAGAAGCACCAAGAGGAGCCCTTCTGTTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db 1321 AAGAAGCACCAAGAGGAGCCCTTCTGTTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy 1381 TGGACGCTGACGCCATCAGCTGCGCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440
Db 1381 TGGACGCTGACGCCATCAGCTGCGCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy 1441 AAGCTGTGGGCAAGCTGAACCTGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCCAG 1500
Db 1441 AAGCTGTGGGCAAGCTGAACCTGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCCAG 1500
Qy 1501 CTGTGCAAGCTGTGCGCGGCGCAAGGCCCTGACCATCTGCGCCCTGACCGAGGAG 1560
Db 1501 CTGTGCAAGCTGTGCGCGGCGCAAGGCCCTGACCATCTGCGCCCTGACCGAGGAG 1560
Qy 1561 GCCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCAGCGGCTGTAC 1620
Db 1561 GCCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCAGCGGCTGTAC 1620
Qy 1621 TACGACCCCGAGAGGACCTGTGGCGGAGATCAGAGAGCGGCGACGACCTGAGACC 1680
Db 1621 TACGACCCCGAGAGGACCTGTGGCGGAGATCAGAGAGCGGCGACGACCTGAGACC 1680
```

```
Qy 1681 TACCAGATCTACGAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATGGC 1740
Db 1681 TACCAGATCTACGAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATGGC 1740
Qy 1741 ACCGCCACCAAGAGGCTGAAGCAGTGAACCGAGGCGGCTGCAGAAATGCCCATGGAG 1800
Db 1741 ACCGCCACCAAGAGGCTGAAGCAGTGAACCGAGGCGGCTGCAGAAATGCCCATGGAG 1800
Qy 1801 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTGCCCATCCAGAAGAGACCTGG 1860
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTGCCCATCCAGAAGAGACCTGG 1860
Qy 1861 GAGACTGTGGTGGACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1920
Db 1861 GAGACTGTGGTGGACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1920
Qy 1921 ACCCCCGCTTGGTGAAGCTGTGTACCGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1980
Db 1921 ACCCCCGCTTGGTGAAGCTGTGTACCGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1980
Qy 1981 ACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGCTG 2040
Db 1981 ACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGCTG 2040
Qy 2041 ACCGACCGGGCCCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Db 2041 ACCGACCGGGCCCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Qy 2101 CTGCAGGCGCATCCAGCTGGCCCTCGAGACAGCGGCGAGGAGTGAACATCGTGACCGAC 2160
Db 2101 CTGCAGGCGCATCCAGCTGGCCCTCGAGACAGCGGCGAGGAGTGAACATCGTGACCGAC 2160
Qy 2161 AGCCAGTACGCGCTGGGCTCATCAGCGCCAGCCCGACAGAGCGAGCGAGCTGGTG 2220
Db 2161 AGCCAGTACGCGCTGGGCTCATCAGCGCCAGCCCGACAGAGCGAGCGAGCTGGTG 2220
Qy 2221 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGGTGCGCCGCC 2280
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGGTGCGCCGCC 2280
Qy 2281 CACAAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGCGATCCGCAAG 2340
Db 2281 CACAAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGCGATCCGCAAG 2340
Qy 2341 GTGCTGTTCTTGGACGCGCATCGATGGCGGCTGATCTACAGTACATGACGACCTG 2400
Db 2341 GTGCTGTTCTTGGACGCGCATCGATGGCGGCTGATCTACAGTACATGACGACCTG 2400
Qy 2401 TACGTGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457
Db 2401 TACGTGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457
```

## RESULT 5

```
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
XX 15-MAY-2002 (first entry)
XX
XX Synthetic construct PR975(+) SEQ ID NO:30.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gsg; pol; vif; vpr; tat; rev; vpu; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
XX Human immunodeficiency virus; type C.
OS
XX Synthetic.
XX
XX WO200204493-A2.
XX
```



PD	17-JAN-2002.	QY	541	GTGCGCCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGCGCATCGGACCGTG	600
XX		DB	541	GTGCGCCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGCGCATCGGACCGTG	600
PF	05-JUL-2001; 2001WO-US021241.				
XX		QY	601	CTGATCGGCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGTGC	660
XX	05-JUL-2000; 2000US-00610313.	DB	601	CTGATCGGCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGTGC	660
PA	(CHIR ) CHIRON CORP.				
PA	(UYST-) UNIV STELLENBOSCH.	QY	661	ACCTGAACTTCCCATCGACCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATG	720
XX	Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	DB	661	ACCTGAACTTCCCATCGACCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATG	720
XX	WPI; 2002-154920/20.	QY	721	GACGGCCCAAGGTGAAGCAGTGGCCCTGACGAGGAGAGATCAAGGCCCTGACCGCC	780
XX		DB	721	GACGGCCCAAGGTGAAGCAGTGGCCCTGACGAGGAGAGATCAAGGCCCTGACCGCC	780
PT	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in	QY	781	ATCTGGAGGAGATGGAGAGGAGGCAAGATCAACAGATCGGCCCGCGAGAACCCCTAC	840
PT	applications including DNA immunization or generation of packaging cell	DB	781	ATCTGGAGGAGATGGAGAGGAGGCAAGATCAACAGATCGGCCCGCGAGAACCCCTAC	840
PT	lines, particularly in gene therapy.	QY	841	AACACCCCGTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	900
XX	Claim 1; Fig 8; 233pp; English.	DB	841	AACACCCCGTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	900
CC	The present invention describes expression cassettes comprising a	QY	901	TTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCAC	960
CC	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV	DB	901	TTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCAC	960
CC	type C polypeptides. The expression cassettes comprise any of the HIV	QY	961	CCCGCGGCTGTAAG	1020
CC	type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef	DB	961	CCCGCGGCTGTAAG	1020
CC	(i). (i) have immunostimulant activity and can be used in gene therapy.	QY	1021	AGCGTCCCTGACGAGGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC	1080
CC	The HIV type C polynucleotides are useful in applications including DNA	DB	1021	AGCGTCCCTGACGAGGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC	1080
CC	immunisation, generation of packaging cell lines, and production of HIV	QY	1081	AACGAG	1140
CC	Type C proteins. The polynucleotides are particularly useful in gene	DB	1081	AACGAG	1140
CC	therapy and DNA immunisation applications. ABI9942 to ABL40054 and	QY	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCTTGAGGCCCTTCCGGCGCCGCAACCCC	1200
CC	ABB06204 to ABB06215 represent sequences used in the exemplification of	DB	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCTTGAGGCCCTTCCGGCGCCGCAACCCC	1200
CC	the present invention	QY	1201	GAGATCGTGTATACCA-----GGCCCGCTGTAGTGGCGAGCGACCTGGAGATCGGC	1254
XX	Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;	DB	1201	GAGATCGTGTATACCA-----GGCCCGCTGTAGTGGCGAGCGACCTGGAGATCGGC	1254
Query Match	99.2%; Score 2442.2; DB 6; Length 2469;	QY	1255	CAGCAGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGGCTTCCACACC	1314
Best Local Similarity	99.6%; Pred. No. 1.5e-294;	DB	1255	CAGCAGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGGCTTCCACACC	1314
Matches 2460; Conservative	0; Mismatches 3; Indels 6; Gaps 1;	QY	1261	CAGCAGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGGCTTCCACACC	1320
		DB	1261	CAGCAGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGGCTTCCACACC	1320
1	GTGACGCCCACTCAAGGGCCCCAAGCGCATCATCAAGTGTTCACATCGCGCAAGAGGGC	QY	1315	CCGCAAGAGAGACCAAG	1374
1	GTGACGCCCACTCAAGGGCCCCAAGCGCATCATCAAGTGTTCACATCGCGCAAGAGGGC	DB	1315	CCGCAAGAGAGACCAAG	1374
61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTTCACATCGCGCAAGAGGGC	QY	1321	CCGCAAGAGAGACCAAG	1380
61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTTCACATCGCGCAAGAGGGC	DB	1321	CCGCAAGAGAGACCAAG	1380
121	CACATCGCGCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG	QY	1375	GACAGTGGAGCTGCGAGCCCATCGAGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAG	1434
121	CACATCGCGCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG	DB	1375	GACAGTGGAGCTGCGAGCCCATCGAGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAG	1434
181	GGCCACAGATGAAGGAGTGCACCGAGCGCCAGGCAACTTCTTCGCGAGGAGCTGGCC	QY	1381	GACAGTGGAGCTGCGAGCCCATCGAGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAG	1440
181	GGCCACAGATGAAGGAGTGCACCGAGCGCCAGGCAACTTCTTCGCGAGGAGCTGGCC	DB	1381	GACAGTGGAGCTGCGAGCCCATCGAGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAG	1440
241	TTCCCGCCAGGAGGCGCGGAGTTCGCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC	QY	1435	ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCGAGATCTACCCCGGATCAAGGTG	1494
241	TTCCCGCCAGGAGGCGCGGAGTTCGCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC	DB	1435	ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCGAGATCTACCCCGGATCAAGGTG	1494
301	AGCCGCGAGCTGCAAGTGGCGGCGCAACCCCGCGAGCGAGGCGCGCGCGCGCGCGAG	QY	1441	ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCGAGATCTACCCCGGATCAAGGTG	1500
301	AGCCGCGAGCTGCAAGTGGCGGCGCAACCCCGCGAGCGAGGCGCGCGCGCGCGAG	DB	1441	ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCGAGATCTACCCCGGATCAAGGTG	1500
361	GGCACCTGAACTTCCCGCGAGATCACCTGTGGCAGCGCCCTGCTGCTGAGCATCAAGGTG	QY	1495	CGCCAGCTGTGCAAGTGTGCGGCGCCCAAGGCCCTGACCGACATCTGTCGCCCTGACC	1554
361	GGCACCTGAACTTCCCGCGAGATCACCTGTGGCAGCGCCCTGCTGCTGAGCATCAAGGTG	DB	1495	CGCCAGCTGTGCAAGTGTGCGGCGCCCAAGGCCCTGACCGACATCTGTCGCCCTGACC	1554
421	GGCGCCAGATCAAGGAGGCGCTGCTGACACCGCGCGCGCGCGCGCGCGCGCGAG	QY	1501	CGCCAGCTGTGCAAGTGTGCGGCGCCCAAGGCCCTGACCGACATCTGTCGCCCTGACC	1560
421	GGCGCCAGATCAAGGAGGCGCTGCTGACACCGCGCGCGCGCGCGCGCGCGAG	DB	1501	CGCCAGCTGTGCAAGTGTGCGGCGCCCAAGGCCCTGACCGACATCTGTCGCCCTGACC	1560
481	ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCAAG	QY	1555	GAGGAGCCGAGCTGGAGCTGGCGAGAACCGCGAGATCTGTCGCCGAGCGCGTGCACGGC	1614
481	ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCAAG	DB	1555	GAGGAGCCGAGCTGGAGCTGGCGAGAACCGCGAGATCTGTCGCCGAGCGCGTGCACGGC	1614
481	ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCAAG	QY	1615	GTGTACTACGACCCCAAGAGGAGCTGGTGGCGCGAGATCCAGAGAGAGAGAGAGAGAG	1674
481	ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCAAG	DB	1615	GTGTACTACGACCCCAAGAGGAGCTGGTGGCGCGAGATCCAGAGAGAGAGAGAGAGAG	1674

```

Db 1621 GTGTTACTAGACCCAGAGGACCTGGTGGCCGAGATCCAGAACGAGGCGCCAGCAGCAG 1680
Qy 1675 TGGACCTTACAGATCTACCCAGAGCCCTTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1734
Db 1681 TGGACCTTACAGATCTACCCAGAGCCCTTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1740
Qy 1735 ATGGCAGCGCCGACACCAACGAGCTGAAGCAGCTGACCGAGGCGCGTGCAGAGATCGCC 1794
Db 1741 ATGGCAGCGCCGACACCAACGAGCTGAAGCAGCTGACCGAGGCGCGTGCAGAGATCGCC 1800
Qy 1795 ATGGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAG 1854
Db 1801 ATGGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAG 1860
Qy 1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCGCACCTGGATCCCCGAGTGGGAGTTC 1914
Db 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCGCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy 1915 GTGAACACCCCCCTCTGGTGAAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGGC 1974
Db 1921 GTGAACACCCCCCTCTGGTGAAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGGC 1980
Qy 1975 GCGGAGACCTTCTAGCTGAGCGGCGCGCCACCGCAGAGCCAGATCGGCAAGGCGGC 2034
Db 1981 GCGGAGACCTTCTAGCTGAGCGGCGCGCCACCGCAGAGCCAGATCGGCAAGGCGGC 2040
Qy 2035 TACGTGACCGACCGGGCGCGCAGAAAGATCTGAGCCTGACCGAGACCAACCAAGAG 2094
Db 2041 TACGTGACCGACCGGGCGCGCAGAAAGATCTGAGCCTGACCGAGACCAACCAAGAG 2100
Qy 2095 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGAGGACACGCGCAGCGAGGTGAACATCGT 2154
Db 2101 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGAGGACACGCGCAGCGAGGTGAACATCGT 2160
Qy 2155 ACCGACGACGATGACCGCTGGGCATCATCCAGGCCCGCAGCGCCAGAGCGAGCGAG 2214
Db 2161 ACCGACGACGATGACCGCTGGGCATCATCCAGGCCCGCAGCGCCAGAGCGAGCGAG 2220
Qy 2215 CTGGTGAACACAGATCATCGACGAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2274
Db 2221 CTGGTGAACACAGATCATCGACGAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2280
Qy 2275 CCGGCCCAAGGGGATCGCGCGCAACGAGCAGATCGAAGCTGTGTAGCAAGGGCATC 2334
Db 2281 CCGGCCCAAGGGGATCGCGCGCAACGAGCAGATCGAAGCTGTGTAGCAAGGGCATC 2340
Qy 2335 CGCAAGGTGCTTCTTGGAGCGGATCGATGGCGGCATCGTATCTACAGTATACGAC 2394
Db 2341 CGCAAGGTGCTTCTTGGAGCGGATCGATGGCGGCATCGTATCTACAGTATACGAC 2400
Qy 2395 GACCTGTACGTGGGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2454
Db 2401 GACCTGTACGTGGGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2460
Qy 2455 GGTGAATTC 2463
Db 2461 GGTGAATTC 2469

```

## RESULT 6

ADM73764

ID ADM73764 standard; DNA; 2469 BP.

XX AC

XX AC ADM73764;

XX DT

XX 03-JUN-2004 (first entry)

XX DE

XX HIV-1 polynucleotide #7.

XX KW

XX HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;

XX RW

XX HIV type C protein; immunostimulant.

XX XX

OS Human immunodeficiency virus 1.

XX PN

XX US2003223961-A1.

XX PD

XX 04-DEC-2003.

XX PF

XX 05-JUL-2001; 2001US-00899575.

XX PR

XX 05-JUL-2000; 2000US-00610313.

XX PA

XX (MEGE/) MEGEDE J Z.

XX PA

XX (BARN/) BARNETT S W.

XX PA

XX (ENGE/) ENGELBRECHT S.

XX PA

XX (RENS/) RENSBURG E J V.

XX XX

PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJ;

XX XX

XX WPI; 2004-060515/06.

XX XX

PS Claim 1; SEQ ID NO 30; 160pp; English.

XX XX

The invention relates to an expression cassette comprising a polynucleotide sequence encoding an HIV pol polypeptide. The invention also relates to a recombinant expression system for use in a host cell comprising an expression cassette, where the polynucleotide sequence further comprises control elements capable of driving expression in the selected host cell, a cell comprising an expression cassette where the polynucleotide sequence further comprises control elements compatible with the expression in the cell and a composition for generating an immunological response, comprising an expression cassette. The expression cassette and the methods of the invention are useful in eliciting an immune response, in DNA immunisation, in generation of packaging cell lines and in producing HIV type C proteins. This sequence represents an HIV-1 polynucleotide of the invention.

SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 99.2%; Score 2442.2; DB 12; Length 2469;

Best Local Similarity 99.6%; Pred. No. 1.5e-294;

Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy 1 GTCCAGCCACCATGGCGGCGCATGAGCGCCACCGAGCCCAACATCTGATGCGAG 60

Db 1 GTCAGCCACCATGGCGGCGCATGAGCGCCACCGAGCCCAACATCTGATGCGAG 60

Qy 61 CGCAGCAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGC 120

Db 61 CGCAGCAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGC 120

Qy 121 CACATCGCCGCAACTGCGGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180

Db 121 CACATCGCCGCAACTGCGGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180

Qy 181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGCGCAACTTCTTCGCGAGGACTGSCC 240

Db 181 GGCCACCAGATGAAGGACTGCACCGAGCGCCAGCGCAACTTCTTCGCGAGGACTGSCC 240

Qy 241 TTCCCCCGAGGCAAGGCGCCGAGTTCCTCCAGCGAGCAGAACCGCGCCCAAGCCCCACC 300

Db 241 TTCCCCCGAGGCAAGGCGCCGAGTTCCTCCAGCGAGCAGAACCGCGCCCAAGCCCCACC 300

Qy 301 AGCCGCGAGTGCAGGTGCGCGGCAAAACCCCGCAGCGAGGCGCGCGCGAGGCGCGAG 360

Db 301 AGCCGCGAGTGCAGGTGCGCGGCAAAACCCCGCAGCGAGGCGCGCGCGAGGCGCGAG 360

Qy 361 GGCACCTGACCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGTGAGCATCAAGGTG 420

Db 361 GGCACCTGACCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGTGAGCATCAAGGTG 420

Qy	421	GGCGCCAGATCAAGGAGGCCCTGCTGGACACCGGCGCGACGACACCGTGCTGGAGGAG	480
Dd	421		
Qy	481	ATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGCGGCTTCATCAAG	540
Dd	481		
Qy	541	GTGCGCAGTACGACCAAGATCCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGACCGTG	600
Dd	541		
Qy	601	CTGATCGGCGCCACCCCGTGGAATCATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Dd	601		
Qy	661	ACCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGAACTGAAAGCCCGGCATG	720
Dd	661		
Qy	721	GACGCGCCCAAGGTGAAGAGGAGGCGCAAGATCACCAAGATCGGCGCGAGAGATCAAGGCCCTGACCGCC	780
Dd	721		
Qy	781	ATCTGCGAGGAGATGGAAGAGGAGGCGAAGATCACCAAGATCGGCGCGAGAGATCAAGGCCCTGACCGCC	840
Dd	781		
Qy	841	AACACCCCGCTGTTCCGCCATCAAGAAAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Dd	841		
Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGCTGGGCGATCCCCAC	960
Dd	901		
Qy	961	CCGCGCGCTGGAAGAGAGAGAGCGTGCACCGTGTGGACCGTGGCGAGCGCTACTTC	1020
Dd	961		
Qy	1021	AGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCCAGCATCAAC	1080
Dd	1021		
Qy	1081	AAAGAGACCCCGGCATCCGTTACAGTAAACCGTGTGCGCGCGAGGCGAGC	1140
Dd	1081		
Qy	1141	CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCGCGCAACCCC	1200
Dd	1141		
Qy	1201	GAGATCGTGATCTACCA-----GGCCCCCTGTGACGTGGGCGAGCAGCCTGGAGATCGGC	1254
Dd	1201		
Qy	1255	CAGACCGCGCAAGATCGAGAGCTCGGCAAGACCTGCTGGCTGGGCGCTTACCAACC	1314
Dd	1255		
Qy	1261	CAGACCGCGCAAGATCGAGGAGCTCGCAAGACCTGCTGGCTGGGCGCTTACCAACC	1320
Dd	1261		
Qy	1315	CCGACAAGAGCAGCAGAGGAGCCCTTCTGTTGGATGGGCTACGAGCTGCACCCC	1374
Dd	1315		
Qy	1375	GACAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAGGAGAGCTGGACCGTGAACGAC	1434
Dd	1375		
Qy	1435	ATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGAGCCAGATCTACCCCGGCATCAAGGTG	1494
Dd	1435		
Qy	1441	ATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGAGGATCTTACCCCGGCATCAAGGTG	1500
Dd	1441		

RESULT 7  
ABL39961  
ID ABL39961 standard; DNA; 2457 BP.  
XX

Qy	1495	CGCAGCTGTGCAAGCTGTGCGCGCGCAAGGCCCTGACCGACATCTGTGCCCTGACC	1554
Dd	1501		
Qy	1555	GAGAGGCCGAGCTGAGCTGGCCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGGC	1614
Dd	1561		
Qy	1615	GTCTACTACGACCCAGCAGGAGACCTGTGGCGCGAGATCCAGAAAGCAGAGGCCACGACCG	1674
Dd	1621		
Qy	1675	TGACCTTACCAAGTCTACCAAGAGCCCTTCAAGAACTGGAAGACCGCAAGTACGCCAAG	1734
Dd	1681		
Qy	1735	ATGCGCACCGCCCAACCAAGAGCTGAGAGCAGCTGACCGAGGCGGTGCAAGAGATCGCC	1794
Dd	1741		
Qy	1795	ATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAG	1854
Dd	1801		
Qy	1855	ACCTGGGAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCAGATCGGAGTTC	1914
Dd	1861		
Qy	1915	GTGAAACACCCCGCTGGTGAAGCTGTGTTACAGCTGGAGAGGAGCCCATCATCGGC	1974
Dd	1921		
Qy	1975	GCCGAGACCTTCTACCTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCG	2034
Dd	1981		
Qy	2035	TAGTGACCCAGACCGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACACCAAG	2094
Dd	2041		
Qy	2095	ACGAGCTGACGGCCATTCAGCTGGCCCTGCGAGACAGCGGCGAGCGGTGAACATCGTG	2154
Dd	2101		
Qy	2155	ACGAGACCGAGTACGCCCTGGGCGCATCTCCAGCGCCAGCCCGACAGAGCGAGCGAG	2214
Dd	2161		
Qy	2215	CTGTTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG	2274
Dd	2221		
Qy	2275	CCGCGCAAGGGCATCGGCGCAACGAGCAGATCGACAAAGCTGTGAGCAAGGGGCATC	2334
Dd	2281		
Qy	2335	CGCAAGTGTCTTCTTGGACGCGCATCGATGGGCGCATCTGATCTACCAAGTACATCGAC	2394
Dd	2341		
Qy	2395	GACCTGTACGTGGGCGAGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCAC	2454
Dd	2401		
Qy	2455	GGTGAATTC 2463	
Dd	2461	GGTGAATTC 2469	



Db 1375 TGGACCGTCGAGCCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1434  
 Qy 1441 AAGCTGTGGGCAAGCTGAAGTGGGCGGAGAGATCTACCCCGCATCAAGGTGGCGCAG 1500  
 Db 1435 AAGCTGTGGGCAAGCTGAAGTGGGCGGAGAGATCTACCCCGCATCAAGGTGGCGCAG 1494  
 Qy 1501 CTGTGAAGCTGTCTGGCGGCGGCAAGGCGCTGACCGACATCTGTGCCCTTGACCGAGGAG 1560  
 Db 1495 CTGTGAAGCTGTCTGGCGGCGGCAAGGCGCTGACCGACATCTGTGCCCTTGACCGAGGAG 1554  
 Qy 1561 GCGAGCTGAGCTGTGGCGGCGGCAAGGCGCTGACCGACATCTGTGCCCTTGACCGAGGAG 1620  
 Db 1555 GCGAGCTGAGCTGTGGCGGCGGCAAGGCGCTGACCGACATCTGTGCCCTTGACCGAGGAG 1614  
 Qy 1621 TAGACCCAGCAGGAGCTGTGGCGGCGGCAAGGCGCTGACCGACATCTGTGCCCTTGACCGAGGAG 1680  
 Db 1615 TAGACCCAGCAGGAGCTGTGGCGGCGGCAAGGCGCTGACCGACATCTGTGCCCTTGACCGAGGAG 1674  
 Qy 1681 TACCAGATCTACAGGAGCGCTTCAAGAACTGGAAGACCGGCAAGTACGGCAAGATCGCG 1740  
 Db 1675 TACCAGATCTACAGGAGCGCTTCAAGAACTGGAAGACCGGCAAGTACGGCAAGATCGCG 1734  
 Qy 1741 ACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAAGAGATCGCCATGGAG 1800  
 Db 1735 ACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAAGAGATCGCCATGGAG 1794  
 Qy 1801 AGCATCTGTATCTGGGCGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1860  
 Db 1795 AGCATCTGTATCTGGGCGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1854  
 Qy 1861 GAGACCTGTGTGGGCGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1920  
 Db 1855 GAGACCTGTGTGGGCGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1914  
 Qy 1921 ACCCCCCCTGTGTGAAGCTGTGTGTAACAGCTGGAAGAGAGCCCATCATCGGCGCGAG 1980  
 Db 1915 ACCCCCCCTGTGTGAAGCTGTGTGTAACAGCTGGAAGAGAGCCCATCATCGGCGCGAG 1974  
 Qy 1981 ACCTTCTACGTGAGCGCGCGGCAAGCGGAGACCAAGATCGGCAAGGCGGCTACGTG 2040  
 Db 1975 ACCTTCTACGTGAGCGCGCGGCAAGCGGAGACCAAGATCGGCAAGGCGGCTACGTG 2034  
 Qy 2041 ACCGACCGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACACCAAGAGACCGAG 2100  
 Db 2035 ACCGACCGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACACCAAGAGACCGAG 2094  
 Qy 2101 CTGACGCGCATTCAGCTGGCCCTGACGAGACGCGGCAAGGAGTGAACATCGTGACCGAC 2160  
 Db 2095 CTGACGCGCATTCAGCTGGCCCTGACGAGACGCGGCAAGGAGTGAACATCGTGACCGAC 2154  
 Qy 2161 AGCAGTACGCGCTGGGCAATCATCCAGGCGGCGGCAAGGAGTGAACATCGTGACCGAG 2220  
 Db 2155 AGCAGTACGCGCTGGGCAATCATCCAGGCGGCGGCAAGGAGTGAACATCGTGACCGAG 2214  
 Qy 2221 AACAGATCATTCAGAGAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCCCGCC 2280  
 Db 2215 AACAGATCATTCAGAGAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCCCGCC 2274  
 Qy 2281 CACAAGGGCATTCGCGCGCAACAGCAGATCGACAAGCTGTGTAGCAAGGGGATCCGCAAG 2340  
 Db 2275 CACAAGGGCATTCGCGCGCAACAGCAGATCGACAAGCTGTGTAGCAAGGGGATCCGCAAG 2334  
 Qy 2341 GTGCTGTCTTGGAGCGGCGCATCGATGGCGGATCGTGATCTACCAAGTACATGGAGCCTG 2400  
 Db 2335 GTGCTGTCTTGGAGCGGCGCATCGATGGCGGATCGTGATCTACCAAGTACATGGAGCCTG 2394  
 Qy 2401 TACGTGGGCGGCGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
 Db 2395 TACGTGGGCGGCGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2454  
 Qy 2461 TTC 2463  
 Db 2455 TTC 2457

RESULT 8  
 ADM73766

ID ADM73766 standard; DNA; 2457 BP.

XX ADM73766;

XX AC

XX 03-JUN-2004 (first entry)

XX HIV-1 polynucleotide #9.

XX HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;

XX HIV type C protein; immunostimulant.

XX Human immunodeficiency virus 1.

XX OS

XX PN

XX 04-DEC-2003.

XX 05-JUL-2001; 2001US-00899575.

XX 05-JUL-2000; 2000US-00610313.

XX (MEGE/) MEGEDE J Z.

XX (BARN/) BARNETT S W.

XX (ENGE/) ENGELBRECHT S.

XX (RENS/) RENSBUURG E J V.

XX Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJ;

XX WPI; 2004-060515/06.

XX Claim 1; SEQ ID NO 32; 160pp; English.

XX The invention relates to an expression cassette comprising a polynucleotide sequence encoding an HIV Pol polypeptide. The invention also relates to a recombinant expression system for use in a host cell comprising an expression cassette, where the polynucleotide sequence further comprises control elements capable of driving expression in the selected host cell, a cell comprising an expression cassette where the polynucleotide sequence further comprises control elements compatible with the expression in the cell and a composition for generating an immunological response, comprising an expression cassette. The expression cassette and the methods of the invention are useful in eliciting an immune response, in DNA immunisation, in generation of packaging cell lines and in producing HIV type C proteins. This sequence represents an HIV-1 polynucleotide of the invention.

XX Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 98.9%; Score 2436.2; DB 12; Length 2457;

Best Local Similarity 99.6%; Pred. No. 8.3e-294;

Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy 1 GTGACGCCACCATGGCGGAGGCCATGAGCAGGCCACCGAGCCCAACATCTCTGATCGAG 60

Db 1 GTGACGCCACCATGGCGGAGGCCATGAGCAGGCCACCGAGCCCAACATCTCTGATCGAG 60

Qy 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGC 120

Db 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGC 120

Qy 121 CACATCGCCCAACTTCGCGGCGCCCCCGCAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180

Db 121 CACATCGCCCAACTTCGCGGCGCCCCCGCAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180





Db 2335 GTGCTGTTCTGGAGCGCATGATGGCGCATCGTGATCTACAGTATACGAGCAACCTG 2394  
 Qy 2401 TACGTGGCGCAGCGCGCCCTAGGATGATTAAGATTTCGGGGCTAGCACCGGTGAA 2460  
 Db 2395 TACGTGGCGCAGCGCGCCCTAGGATGATTAAGATTTCGGGGCTAGCACCGGTGAA 2454  
 Qy 2461 TTC 2463  
 Db 2455 TTC 2457  
 RESULT 9  
 ID ACA03548 standard; DNA; 2457 BP.  
 AC ACA03548;  
 XX  
 DT 22-MAY-2003 (first entry)  
 XX  
 DE Synthetic DNA encoding immunogenic HIV peptide #31.  
 XX  
 KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN W02003004657-A1.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021421.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Zur Megede J, Barnett SW, Lian Y;  
 XX  
 DR WPI; 2003-221602/21.  
 XX  
 PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX  
 PS Example 1; Fig 36; 262pp; English.  
 XX  
 CC The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide  
 XX  
 SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;  
 Query Match 98.78; Score 2430.2; DB 8; Length 2457;  
 Best Local Similarity 99.68; Pred. No. 4.6e-293;  
 Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
 Qy 7 GCCACCATGGCCGAGGCGCATGAGCCAGGCCACCAAGCCCATCTGATGCGAGCGCAGC 66  
 |||||

Db 1 GCCACCATGGCCGAGGCGCATGAGCCAGGCCACCAAGCCCATCTCTGATGCGAGCGCAGC 60  
 Qy 67 AACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGGAGGGCCACATC 126  
 |||||  
 Db 61 AACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGGAGGGCCACATC 120  
 |||||  
 Qy 127 GCCCGCAACTGCGCGCCCCCGCAAGAGGGGTGCTTGGAAAGTTCGCGCAAGAGAGGGCCAC 186  
 |||||  
 Db 121 GCCCGCAACTGCGCGCCCCCGCAAGAGGGGTGCTTGGAAAGTTCGCGCAAGAGAGGGCCAC 180  
 |||||  
 Qy 187 CAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGGCTTCCCC 246  
 |||||  
 Db 181 CAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGGCTTCCCC 240  
 |||||  
 Qy 247 CAGGGCAAGCGCGCGAGTTCCCGAGCGAGCAACCGCGCCAGCGCCAGCGCCACAGCGCGC 306  
 |||||  
 Db 241 CAGGGCAAGCGCGCGAGTTCCCGAGCGAGCAACCGCGCCAGCGCCACAGCGCGC 300  
 |||||  
 Qy 307 GAGCTGCAGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 366  
 |||||  
 Db 301 GAGCTGCAGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
 |||||  
 Qy 367 CTGAATTTCCCGCAGATCACTTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCGC 426  
 |||||  
 Db 361 CTGAATTTCCCGCAGATCACTTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCGC 420  
 |||||  
 Qy 427 CAGATCAAGGAGGCGCTGTGGACACCGCGCGCGAGCACACCGTGTGGAGGAGATGAGC 486  
 |||||  
 Db 421 CAGATCAAGGAGGCGCTGTGGACACCGCGCGCGAGCACACCGTGTGGAGGAGATGAGC 480  
 |||||  
 Qy 487 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGCGGGCATCGCGGGCTTTCATCAAGGTGCGC 546  
 |||||  
 Db 481 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGCGGGCATCGCGGGCTTTCATCAAGGTGCGC 540  
 |||||  
 Qy 547 CAGTACGACAGATCTGTGATCGAGATCTGGCGCAAGAGGCCATCGGCACTGCTGTGATC 606  
 |||||  
 Db 541 CAGTACGACAGATCTGTGATCGAGATCTGGCGCAAGAGGCCATCGGCACTGCTGTGATC 600  
 |||||  
 Qy 607 GGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666  
 |||||  
 Db 601 GGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660  
 |||||  
 Qy 667 AACTTTCCTCCATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGACGGC 726  
 |||||  
 Db 661 AACTTTCCTCCATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGACGGC 720  
 |||||  
 Qy 727 CCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGC 786  
 |||||  
 Db 721 CCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGC 780  
 |||||  
 Qy 787 GAGGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTACAACACC 846  
 |||||  
 Db 781 GAGGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTACAACACC 840  
 |||||  
 Qy 847 CCGTGTTCCTCCATCAAGAGAGAGGAGCAGCAACCAAGTGGCGCAAGCTGGTGGATTTCCGC 906  
 |||||  
 Db 841 CCGTGTTCCTCCATCAAGAGAGAGGAGCAGCAACCAAGTGGCGCAAGCTGGTGGATTTCCGC 900  
 |||||  
 Qy 907 GAGCTGAACAGGCGACCCAGGACTTCTTGGAGGAGTGGAGTGGGATCCCCACCCCGCC 966  
 |||||  
 Db 901 GAGCTGAACAGGCGACCCAGGACTTCTTGGAGGAGTGGAGTGGGATCCCCACCCCGCC 960  
 |||||  
 Qy 967 GGCTTGAAGAGAGAGAGGAGCGCTGACCGTCTGAGAGTGGGCGACGCGCTACTTTCAGGCTG 1026  
 |||||  
 Db 961 GGCTTGAAGAGAGAGAGGAGCGCTGACCGTCTGAGAGTGGGCGACGCGCTACTTTCAGGCTG 1020  
 |||||  
 Qy 1027 CCCTTGGACGAGGACTTTCGCAAGATCACCGCCCTTCCATCCCGAGCATCAACAACGAG 1086  
 |||||  
 Db 1021 CCCTTGGACGAGGACTTTCGCAAGATCACCGCCCTTCCATCCCGAGCATCAACAACGAG 1080  
 |||||  
 Qy 1087 ACCCCCGGATCTCGCTTACGATACAGTGTCTGCCCCCGAGGCTGGAGGGGAGGGCCCCAGC 1146  
 |||||  
 Db 1081 ACCCCCGGATCTCGCTTACGATACAGTGTCTGCCCCCGAGGCTGGAGGGGAGGGCCCCAGC 1140  
 |||||

QY 1147 ATCTTCAGAGCAGATGACCAAGATCTGTGAGCCCTTCGCGCCGCGCAACCCCGAGATC 1206  
 Db 1141 ATCTTCAGAGCAGATGACCAAGATCTGTGAGCCCTTCGCGCCGCGCAACCCCGAGATC 1200  
 QY 1207 GTGATCTACCA-----GSCCCCTCTGTACGTGGGCGAGCGACTGGAGATCGGCCAGCAGC 1260  
 Db 1201 GTGATCTACCAAGATGACGACCTGTACGTGGGCGAGCGACTGGAGATCGGCCAGCAGC 1260  
 QY 1261 CGCGCAAGATCAGAGAGCTGCGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGGAC 1320  
 Db 1261 CGCGCAAGATCAGAGAGCTGCGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGGAC 1320  
 QY 1321 AAGAGCAGCAGAGAGCCCTTCTGTGGATGGGCTACGAGCTGCAACCCCGCAAG 1380  
 Db 1321 AAGAGCAGCAGAGAGCCCTTCTGTGGATGGGCTACGAGCTGCAACCCCGCAAG 1380  
 QY 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAG 1440  
 Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAG 1440  
 QY 1441 AAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGGATCAAGGTGCGCCAG 1500  
 Db 1441 AAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGGATCAAGGTGCGCCAG 1500  
 QY 1501 CTGTGCAAGCTGCTGCGGGGCGCCAAAGCCCTGACCGACATCGTGCCTTGAACGAGGAG 1560  
 Db 1501 CTGTGCAAGCTGCTGCGGGGCGCCAAAGCCCTGACCGACATCGTGCCTTGAACGAGGAG 1560  
 QY 1561 GCCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGGAGCCCGTGCAGCGCGTGTAC 1620  
 Db 1561 GCCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGGAGCCCGTGCAGCGCGTGTAC 1620  
 QY 1621 TACGACCCCGCAGAGACCTGTGTGCGGAGATCCAGAAAGCAGGGCCAGCAGCAGTGAGCC 1680  
 Db 1621 TACGACCCCGCAGAGACCTGTGTGCGGAGATCCAGAAAGCAGGGCCAGCAGCAGTGAGCC 1680  
 QY 1681 TACGAGATCTACAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGC 1740  
 Db 1681 TACGAGATCTACAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGC 1740  
 QY 1741 ACCGCCACACCCAGCTGAAGCAGCTGACCGAGCGCGTGCAGAAAGTGCACATGCGCATGAG 1800  
 Db 1741 ACCGCCACACCCAGCTGAAGCAGCTGACCGAGCGCGTGCAGAAAGTGCACATGCGCATGAG 1800  
 QY 1801 AGCATCTGTGATCTGGGCGAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGGAGACCTGG 1860  
 Db 1801 AGCATCTGTGATCTGGGCGAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGGAGACCTGG 1860  
 QY 1861 GAGACCTGTGAGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTGTTGAAC 1920  
 Db 1861 GAGACCTGTGAGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTGTTGAAC 1920  
 QY 1921 ACCCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGCGCCGAG 1980  
 Db 1921 ACCCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGCGCCGAG 1980  
 QY 1981 ACTTCTACGTGAGCGCGCCCAACCCGAGACCAAGATCGGCAAGCCGCGCTACGTG 2040  
 Db 1981 ACTTCTACGTGAGCGCGCCCAACCCGAGACCAAGATCGGCAAGCCGCGCTACGTG 2040  
 QY 2041 ACCGACCGGGCGCGCAGAAGATCGTAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
 Db 2041 ACCGACCGGGCGCGCAGAAGATCGTAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
 QY 2101 CTGACGCGCATTCAGCTGGCCCTGACGACAGCGCAGCGAGTGAACATGTTACCGAC 2160  
 Db 2101 CTGACGCGCATTCAGCTGGCCCTGACGACAGCGCAGCGAGTGAACATGTTACCGAC 2160  
 QY 2161 AGCCAGTACGCTGGGCTATCATCCAGGCCCGAGCCCGAAGAGCGAGGAGCTGTG 2220  
 Db 2161 AGCCAGTACGCTGGGCTATCATCCAGGCCCGAGCCCGAAGAGCGAGGAGCTGTG 2220

QY 2221 AACAGATCATCGACAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGGTGCCCGCC 2280  
 Db 2221 AACAGATCATCGACAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGGTGCCCGCC 2280  
 QY 2281 CACAAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGGCAAGGCGCATCCGCAAG 2340  
 Db 2281 CACAAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGGCAAGGCGCATCCGCAAG 2340  
 QY 2341 GTGCTGTTCTTGGAGCGGCATCGATGCGCGCATCGTGTATCTACCAAGTACATGACGACCTG 2400  
 Db 2341 GTGCTGTTCTTGGAGCGGCATCGATGCGCGCATCGTGTATCTACCAAGTACATGACGACCTG 2400  
 QY 2401 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457  
 Db 2401 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 10  
 ADCL3266  
 ID ADCL3266 standard; DNA; 2457 BP.  
 AC AC  
 XX ADCL3266;  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE DNA of HIV construct p2Pol-opt\_C SEQ ID NO 45.  
 KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX Human immunodeficiency virus.  
 OS  
 FN WO2003004620-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021420.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (UYST-) UNIV STELLENBOSCH.  
 XX  
 PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 XX WPI; 2003-221593/21.  
 XX  
 PS New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.  
 XX  
 XX Disclosure; Fig 42; 301pp; English.  
 XX  
 CC The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intracutaneously, intravenously,  
 CC subcutaneously, intradermally, transdermally, intranasally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C-related sequence of the invention.  
 XX  
 SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 98.7%; Score 2430.2; DB 10; Length 2457;  
 Best Local Similarity 99.6%; Pred. No. 4.6e-293;  
 Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 7 GCACCATGGCGGAGGCGCATGAGCCAGGCGCACAGCGCCCAACATCTCTGATGACGCGCAGC 66  
DB |||||  
1 GCACCATGGCGGAGGCGCATGAGCCAGGCGCACAGCGCCCAACATCTCTGATGACGCGCAGC 60  
QY 67 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCAATC 126  
DB |||||  
61 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCAATC 120  
QY 127 GCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGTGGAGTGGCGCAAGAGGGGCCAC 186  
DB |||||  
121 GCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGTGGAGTGGCGCAAGAGGGGCCAC 180  
QY 187 CAGATGAAGGACTGCACCGAGCGCGCAGCCAACTTCTTCGCGAGGACCTTGGCCTTCCCC 246  
DB |||||  
181 CAGATGAAGGACTGCACCGAGCGCGCAGCCAACTTCTTCGCGAGGACCTTGGCCTTCCCC 240  
QY 247 CAGGGCAAGGCCCGCGAGTTTCCCGAGCGAGAGAAACCGCGCCCAACAGCCCCCAACAGCCGC 306  
DB |||||  
241 CAGGGCAAGGCCCGCGAGTTTCCCGAGCGAGAGAAACCGCGCCCAACAGCCCCCAACAGCCGC 300  
QY 307 GAGCTGAGGTGCGCGCGGCAACCCCCCGAGGAGCGCGCGCGCGAGCGCCAGGGGCACC 366  
DB |||||  
301 GAGCTGAGGTGCGCGCGGCAACCCCCCGAGGAGCGCGCGCGCGAGCGCCAGGGGCACC 360  
QY 367 CTGAACCTTCCCGCAGATCACCTGTGTGCGAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 426  
DB |||||  
361 CTGAACCTTCCCGCAGATCACCTGTGTGCGAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 420  
QY 427 CAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGACACCGTGTGTGGAGGAGATGAGC 486  
DB |||||  
421 CAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGACACCGTGTGTGGAGGAGATGAGC 480  
QY 487 CTGCCCGGCAAGTGGAAAGCCAGATGATCGCGGGCATCGCGCGCTTCATCAAGGTGCGC 546  
DB |||||  
481 CTGCCCGGCAAGTGGAAAGCCAGATGATCGCGGGCATCGCGCGCTTCATCAAGGTGCGC 540  
QY 547 CAGTACCAACAGATCTCTGATCGAGATCTGGCGGAAGAGGCATCTGGCACCGTGTCTGATC 606  
DB |||||  
541 CAGTACCAACAGATCTCTGATCGAGATCTGGCGGAAGAGGCATCTGGCACCGTGTCTGATC 600  
QY 607 GGCCCCACCCCGTGAACATCATGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG 666  
DB |||||  
601 GGCCCCACCCCGTGAACATCATGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG 660  
QY 667 AACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGGAACCGC 726  
DB |||||  
661 AACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGGAACCGC 720  
QY 727 CCAAGGTGAAGCAGTGGCCCTTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
DB |||||  
721 CCAAGGTGAAGCAGTGGCCCTTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGC 780  
QY 787 GAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCGCCGAGAACCCCTTACAAACC 846  
DB |||||  
781 GAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCGCCGAGAACCCCTTACAAACC 840  
QY 847 CCGTGTTCGCATCAAGAGAGGAGCAGCACCAAGTGGCGCAGCTGGTGGACTTCCGC 906  
DB |||||  
841 CCGTGTTCGCATCAAGAGAGGAGCAGCACCAAGTGGCGCAGCTGGTGGACTTCCGC 900  
QY 907 GAGCTGAACAGGCGACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCGCCACCCCGCC 966  
DB |||||  
901 GAGCTGAACAGGCGACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCGCCACCCCGCC 960  
QY 967 GGCTGAAGAGAGAGAGCGTGACCGTGTGGAGTGGCGCGACCGCTTCTTCAGCGTG 1026  
DB |||||  
961 GGCTGAAGAGAGAGAGCGTGACCGTGTGGAGTGGCGCGACCGCTTCTTCAGCGTG 1020  
QY 1027 CCGCTGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGACCATCAACACGAG 1086  
DB |||||  
1021 CCGCTGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGACCATCAACACGAG 1080

QY 1087 ACCCCGCGCATCCGCTACCAAGTACAAAGTGTCTCCCGAGGGCTGGAAAGGCGAGCCCCCAGC 1146  
DB |||||  
1081 ACCCCGCGCATCCGCTACCAAGTACAAAGTGTCTCCCGAGGGCTGGAAGGGCAGCCCCCAGC 1140  
QY 1147 ATCTTCCAGAGCAGCATGACCAAGATCTTGAGGCCCTTCCGCGCCCGCAACCCCGAGATC 1206  
DB |||||  
1141 ATCTTCCAGAGCAGCATGACCAAGATCTTGAGGCCCTTCCGCGCCCGCAACCCCGAGATC 1200  
QY 1207 GTGATCTTACCA-----GGCCCCCTGTAGTGGGAGCGACCTTGGAGATCGGCCAGCAC 1260  
DB |||||  
1201 GTGATCTTACCAAGTACATGGACGACCTGTACTGGGAGCGACCTTGGAGATCGGCCAGCAC 1260  
QY 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTCTGCGCTGGGGCTTCAACACCCCGAC 1320  
DB |||||  
1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCACCTCTGCGCTGGGGCTTCAACACCCCGAC 1320  
QY 1321 AAGAAGCACAGAAAGAGCCCCCTTCTGTGTGATGGGCTACGAGCTGCAACCCCGACAAG 1380  
DB |||||  
1321 AAGAAGCACAGAAAGAGCCCCCTTCTGTGTGATGGGCTACGAGCTGCAACCCCGACAAG 1380  
QY 1381 TGGACCGTGCAGCCCATCGAGCTGCGCGAGAAAGAGAGCTGGACCGTGAACGACATCCAG 1440  
DB |||||  
1381 TGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440  
QY 1441 AAGCTGTGGGCAAGCTGAACTGTGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500  
DB |||||  
1501 CTGTGCAAGCTGTGCGCGCGCCAAAGCCCTGACCGACATCTGTGCCCTTGAACCGAGGAG 1560  
DB |||||  
1501 CTGTGCAAGCTGTGCGCGCGCCAAAGCCCTGACCGACATCTGTGCCCTTGAACCGAGGAG 1560  
QY 1561 GCCGAGCTGAGCTGCGCGGAGAACCGCGAGATCTCTGCGGAGCCCGTGACCGGCTGTAC 1620  
DB |||||  
1561 GCCGAGCTGAGCTGCGCGGAGAACCGCGAGATCTCTGCGGAGCCCGTGACCGGCTGTAC 1620  
QY 1621 TACGACCCCAAGCAAGCACTGGTGGCCGAGATCCAGAAGCAGAGGCCACAGCCAGTGAACC 1680  
DB |||||  
1621 TACGACCCCAAGCAAGCACTGGTGGCCGAGATCCAGAAGCAGAGGCCACAGCCAGTGAACC 1680  
QY 1681 TACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAGCCCGGCAAGTACGCCAAGATCGGC 1740  
DB |||||  
1681 TACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAGAACCCGGCAAGTACGCCAAGATCGGC 1740  
QY 1741 ACCGCCACACCAACCAAGCTGAAGCAGCTGACCGAGGCCGTGAGAAGAGATCGCCATGGAG 1800  
DB |||||  
1741 ACCGCCACACCAACCAAGCTGAAGCAGCTGACCGAGGCCGTGAGAAGAGATCGCCATGGAG 1800  
QY 1801 AGCATCGTGTCTGGGCAAGACCCCAAGTTCCGCTGCCCCATCCAGAAGAGGAGCCTGG 1860  
DB |||||  
1801 AGCATCGTGTCTGGGCAAGACCCCAAGTTCCGCTGCCCCATCCAGAAGAGGAGCCTGG 1860  
QY 1861 GAGACCTGTGGACCGGACTACTGCGAGGCCACCTTGGATCCCGAGTGGGAGTTCTGTGAAC 1920  
DB |||||  
1861 GAGACCTGTGTGGACCGGACTACTGCGAGGCCACCTTGGATCCCGAGTGGGAGTTCTGTGAAC 1920  
QY 1921 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCTCGGCGCCGAG 1980  
DB |||||  
1921 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCTCGGCGCCGAG 1980  
QY 1981 ACCTTCTACCTGGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTTG 2040  
DB |||||  
1981 ACCTTCTACCTGGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTTG 2040  
QY 2041 ACCGACCGGGGCCGCGAAGATCTGTAGCCTGACCGAGACCAACCAAGAGACCCGAG 2100  
DB |||||  
2041 ACCGACCGGGGCCGCGAAGATCTGTAGCCTGACCGAGACCAACCAAGAGACCCGAG 2100  
QY 2101 CTGCAAGGCAATCCAGCTGGGCCCTTGCAGGACGCGGAGCGAGGTGAACATCTGTGACCGAC 2160  
DB |||||  
2101 CTGCAAGGCAATCCAGCTGGGCCCTTGCAGGACGCGGAGCGAGGTGAACATCTGTGACCGAC 2160  
QY 2161 AGCCAGTACGCCCCCTGGGCAATCATCCAGGCCCAAGCCCGAAGAGCGGAGAGCGAGCTGGT 2220

Db 2161 AGCCAGTACGCCCTGGGATCATCATCAGGCCCGACAGAGGAGGAGGAGTGGTG 2220  
 Qy 2221 AACACAGATCATGACGAGCTGATCAAGAGAGAGAGTGTACTCTGAGCTGGGTGCCGCC 2280  
 Db 2221 AACACAGATCATGACGAGCTGATCAAGAGAGAGAGTGTACTCTGAGCTGGGTGCCGCC 2280  
 Qy 2281 CACAAGGCGATCGGCGGCAACAGAGATCGAAGAGTGTGTAGCAAGGCGATCGCAAG 2340  
 Db 2281 CACAAGGCGATCGGCGGCAACAGAGATCGAAGAGTGTGTAGCAAGGCGATCGCAAG 2340  
 Qy 2341 GTGCTGTTCTGGAGCGGATCGATGCGGCGATCGTGATCTACCACTACATGAGACCTG 2400  
 Db 2341 GTGCTGTTCTGGAGCGGATCGATGCGGCGATCGTGATCTACCACTACATGAGACCTG 2400  
 Qy 2401 TACGTGGGAGCGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457  
 Db 2401 TACGTGGGAGCGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 11

ACA03546  
 ID ACA03546 standard; DNA; 2445 BP.  
 AC ACA03546;  
 XX  
 DT 22-MAY-2003 (first entry)  
 XX  
 DE Synthetic DNA encoding immunogenic HIV peptide #29.  
 XX  
 KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003004657-A1.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021421.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Zur Megede J, Barnett SW, Lian Y;  
 XX  
 DR WPI; 2003-221602/21.  
 XX  
 PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX  
 PS Example 1; Fig 34; 262pp; English.  
 XX  
 CC The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide

XX SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;  
 Query Match  
 Best Local Similarity 98.4%; Score 2422.6; DB 8; Length 2445;  
 Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;  
 Qy 7 GCCACCATGGCCGAGGCGCATGAGCCAGCCAGCGCAGCATCTCTGATGAGGCGCAGC 66  
 Db 1 GCCACCATGGCCGAGGCGCATGAGCCAGCCAGCGCAGCATCTCTGATGAGGCGCAGC 60  
 Qy 67 AACCTCAAGGCGCCAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGCCACATC 126  
 Db 61 AACCTCAAGGCGCCAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGCCACATC 120  
 Qy 127 GCCCGCAACTCCCGCGCCCGCAAGAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCCAC 186  
 Db 121 GCCCGCAACTCCCGCGCCCGCAAGAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCCAC 180  
 Qy 187 CAGATGAAGGACTGCAACGAGCGCCAGCCAACTTTCTCCGCGAGGAGCTGGCCTTCCCC 246  
 Db 181 CAGATGAAGGACTGCAACGAGCGCCAGCCAACTTTCTCCGCGAGGAGCTGGCCTTCCCC 240  
 Qy 247 CAGGCGAAGGCGCCGAGTTCCTCCAGCGAGCAGAACCGCGCCACAGCCCCACAGCGCG 306  
 Db 241 CAGGCGAAGGCGCCGAGTTCCTCCAGCGAGCAGAACCGCGCCACAGCCCCACAGCGCG 300  
 Qy 307 GAGCTGCAAGTGGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 366  
 Db 301 GAGCTGCAAGTGGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
 Qy 367 CTGAATTTCCCGCAGATCACCTGTGTGCGAGCGCCCTCTGTGTGAGCATCAAGTGGCGCG 426  
 Db 361 CTGAATTTCCCGCAGATCACCTGTGTGCGAGCGCCCTCTGTGTGAGCATCAAGTGGCGCG 420  
 Qy 427 CAGATCAAGGAGGCGCTGTGACACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 486  
 Db 421 CAGATCAAGGAGGCGCTGTGCGCACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 480  
 Qy 487 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCTCATCAAGTGGCG 546  
 Db 481 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCTCATCAAGTGGCG 540  
 Qy 547 CAGTACACCAAGATCTCTGATCGAGATCTGCGCGCAAGAGCGCATCGGCACCGTGTGATC 606  
 Db 541 CAGTACACCAAGATCTCTGATCGAGATCTGCGCGCAAGAGCGCATCGGCACCGTGTGATC 600  
 Qy 607 GCGCCCGACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666  
 Db 601 GCGCCCGACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660  
 Qy 667 AACTTTCCCGCATCAGCCCGCATCGAGACCGTGGCGGAGCTGAAGCGCGCATGAGCGCGC 726  
 Db 661 AACTTTCCCGCATCAGCCCGCATCGAGACCGTGGCGGAGCTGAAGCGCGCATGAGCGCGC 720  
 Qy 727 CCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGC 786  
 Db 721 CCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGC 780  
 Qy 787 GAGGAGATGGAAGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACAGACC 846  
 Db 781 GAGGAGATGGAAGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACAGACC 840  
 Qy 847 CCGGTGTTCCCGCATCAAGAGAGAGCAGCACCAAGTGGCGCGCAAGCTGGTGCATCTCGC 906  
 Db 841 CCGGTGTTCCCGCATCAAGAGAGAGCAGCACCAAGTGGCGCGCAAGCTGGTGCATCTCGC 900  
 Qy 907 GAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGCATCCCCACCCCGCC 966  
 Db 901 GAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGCATCCCCACCCCGCC 960  
 Qy 967 GGCCTGAAGAGAGAGAGCGTGCACCGTGTGAGCGTGGCGCGAGCGCTTACTTCAGCGTG 1026

Db 961 GGCCTGAAGAAGAGAGCGTGTGACCGTGTCTGGAGCGTGGCGGACGCTACTTTCAGCGTG 1020  
Qy 1027 CCCCTGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCGACATCAACAACGAG 1086  
Db 1021 CCCCTGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCGACATCAACAACGAG 1080  
Qy 1087 ACCCCCGGCACTCCGCTACCAAGTACAAAGTGTCTGCCCGAGGCTGGAAGGGGAGCGCCGAGC 1146  
Db 1081 ACCCCCGGCACTCCGCTACCAAGTACAAAGTGTCTGCCCGAGGCTGGAAGGGGAGCGCCGAGC 1140  
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGTAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGTAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACCGAGCCCGCTGTAGTGGGACAGCGACCTGTGAGATCGGCGACGACCGCGCC 1266  
Db 1201 GTGATCTACCGAGCCCGCTGTAGTGGGACAGCGACCTGTGAGATCGGCGACGACCGCGCC 1260  
Qy 1267 AAGATCGAGGAGCTGCGCAAGCAGCTGTGGCTGGGGCTTACCAACCCCGCAAGAAAG 1326  
Db 1261 AAGATCGAGGAGCTGCGCAAGCAGCTGTGGCTGGGGCTTACCAACCCCGCAAGAAAG 1320  
Qy 1327 CACCAGAGGAGCCCCCTTCTGTGTGATGGGTACAGAGCTGCACCCCGCAAGAGTGAC 1386  
Db 1321 CACCAGAGGAGCCCCCTTCTGTGTGATGGGTACAGAGCTGCACCCCGCAAGAGTGAC 1374  
Qy 1387 GTGAGCCCATCGAGCTGCCCGAGAGGAGCTGGACCGTGGACGATCCAGAGCTG 1446  
Db 1375 GTGAGCCCATCGAGCTGCCCGAGAGGAGCTGGACCGTGGACGATCCAGAGCTG 1434  
Qy 1447 GTGGGCAAGCTGAATCTGGGCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGC 1506  
Db 1435 GTGGGCAAGCTGAATCTGGGCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGC 1494  
Qy 1507 AAGTGTGTGGCGCGCCCAAGGCCCTGACCGACATCTGTGCCCTTGAACCGAGGAGCCGAG 1566  
Db 1495 AAGTGTGTGGCGCGCCCAAGGCCCTGACCGACATCTGTGCCCTTGAACCGAGGAGCCGAG 1554  
Qy 1567 CTGAGCTGGCGGAGAACCGGAGATCTGGCGAGCGCGTGTGACGGGCTGTACTAGAC 1626  
Db 1555 CTGAGCTGGCGGAGAACCGGAGATCTGGCGAGCGCGTGTGACGGGCTGTACTAGAC 1614  
Qy 1627 CCAGAGGAGACCTGTGGCGGAGATCCAGAGGAGCGGCACACGACAGTGGACCTACAG 1686  
Db 1615 CCAGAGGAGACCTGTGGCGGAGATCCAGAGGAGCGGCACACGACAGTGGACCTACAG 1674  
Qy 1687 ATCTACAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1746  
Db 1675 ATCTACAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1734  
Qy 1747 CACCAACGACGTGAAGCAGCTGACCGAGCGCGTGTGAGAGATCGGCATGGAGGATC 1806  
Db 1735 CACCAACGACGTGAAGCAGCTGACCGAGCGCGTGTGAGAGATCGGCATGGAGGATC 1794  
Qy 1807 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGAC 1866  
Db 1795 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGAC 1854  
Qy 1867 TGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1926  
Db 1855 TGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1914  
Qy 1927 CCCCTGTGAGCTGTGTACCACTGGAGGAGGCCCATCATCGGCGCGGAGACCTTC 1986  
Db 1915 CCCCTGTGAGCTGTGTACCACTGGAGAGGAGGCCCATCATCGGCGCGGAGACCTTC 1974  
Qy 1987 TACGTGACGCGCGCGCCCAACCGGAGACCAAGATCGCGAAGCGCGCTACGTGACCGAC 2046  
Db 1975 TACGTGACGCGCGCGCCCAACCGGAGACCAAGATCGCGAAGCGCGCTACGTGACCGAC 2034  
Qy 2047 CGGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGAG 2106  
Db 2035 CGGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGAG 2094

Qy 2107 GCATCCAGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTGACCGACAGCCAG 2166  
Db 2095 GCATCCAGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTGACCGACAGCCAG 2154  
Qy 2167 TACGCCCTGGGCACTATCCAGGCCAGCCCGCAAGAGCGAGCGAGCTGGTGAACAG 2226  
Db 2155 TACGCCCTGGGCACTATCCAGGCCAGCCCGCAAGAGCGAGCGAGCTGGTGAACAG 2214  
Qy 2227 ATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGGTGCCGCCACAAG 2286  
Db 2215 ATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGGTGCCGCCACAAG 2274  
Qy 2287 GGATCGGCGCGCAACGAGCAGATCGCAACGCTGTGTAGCAAGGCGCATCCGCAAGGTGCTG 2346  
Db 2275 GGATCGGCGCGCAACGAGCAGATCGCAACGCTGTGTAGCAAGGCGCATCCGCAAGGTGCTG 2334  
Qy 2347 TTCTTGGACGGCATCGATGCGCGCATCGTGTATCTACAGTACATGGACGACCTGTACGTG 2406  
Db 2335 TTCTTGGACGGCATCGATGCGCGCATCGTGTATCTACAGTACATGGACGACCTGTACGTG 2394  
Qy 2407 GGCAGCGGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457  
Db 2395 GGCAGCGGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2445

## RESULT 12

ADC13264

ID ADC13264 standard; DNA; 2445 BP.

XX AC ADC13264;

XX AC ADC13264;

DT 18-DEC-2003 (first entry)

XX AC ADC13264;

DE DNA of HIV construct p2Pol-opt-YMWM\_C SEQ ID NO 43.

XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;

XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX Human immunodeficiency virus.

XX WO2003004620-A2.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.

XX 31-AUG-2001; 2001US-0316860P.

XX 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2003-221593/21.

XX New expression cassette comprising a polynucleotide sequence encoding a

XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

XX Prot, or Rev polypeptide, useful for immunization, or generating

XX packaging cell lines.

XX Disclosure; Fig 40; 301pp; English.

XX The invention relates to a novel expression cassette comprising a

XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,

XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel

XX expression cassette can be used to treat HIV type C by gene therapy or

XX used in the development of a vaccine. The gene delivery vector is

XX administered intramuscularly, intramucosally, intranasally,

XX subcutaneously, intradermally, transdermally, intravaginally,

XX intrarectally, orally or intravenously. The expression cassette is useful

CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.

SX Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 98.4%; Score 2422.6; DB 10; Length 2445;  
 Best Local Similarity 99.6%; Pred. No. 4e-292;  
 Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy	7	GCACATATGCGCGAGCCATGAGCCAGCCAGCGCCCAACATCCTGTATGAGCGGACG 66
Db	1	GCACATATGCGCGAGCCATGAGCCAGCGCCCAACATCCTGTATGAGCGGACG 60
Qy	67	AATTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGCCACATC 126
Db	61	AATTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGCCACATC 120
Qy	127	GCCGCAACTGCGCGCGCCCCCGGCAAGAGGGCTGCTGGAATGCGGCAAGAGGGCCAC 186
Db	121	GCCGCAACTGCGCGCGCCCCCGGCAAGAGGGCTGCTGGAATGCGGCAAGAGGGCCAC 180
Qy	187	CAGATGAAGGACTGCAACGAGCGCCAGCGCAACTTCTTCGCGAGGACCTGSCCTTTCC 246
Db	181	CAGATGAAGGACTGCAACGAGCGCCAGCGCAACTTCTTCGCGAGGACCTGSCCTTTCC 240
Qy	247	CAGGCAAGGGCCCGAGTTCCCAAGCGAGCAGAAACCGCGCAACAGCCCCACAGCGCG 306
Db	241	CAGGCAAGGGCCCGAGTTCCCAAGCGAGCAGAAACCGCGCAACAGCCCCACAGCGCG 300
Qy	307	GAGTGCAGGTGCGCGGCAACCCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db	301	GAGTGCAGGTGCGCGGCAACCCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Qy	367	CTGAATTTCCCGCAGATCACCTGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
Db	361	CTGAATTTCCCGCAGATCACCTGTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Qy	427	CAGATCAAGGAGGCGCTCTGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486
Db	421	CAGATCAAGGAGGCGCTCTGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Qy	487	CTGCGCGCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGCGCTTCAATCAAGTGC 546
Db	481	CTGCGCGCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGCGCTTCAATCAAGTGC 540
Qy	547	CAGTACGACAGATCTCATCGAGATCTGCGCAGAGAGGCGCATCGGCACCGTCTGATC 606
Db	541	CAGTACGACAGATCTCATCGAGATCTGCGCAGAGAGGCGCATCGGCACCGTCTGATC 600
Qy	607	GCGCCCAACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTG 666
Db	601	GCGCCCAACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTG 660
Qy	667	AATTTCCCATCAGCCCCCATCGAGACCGTGGCGTGAAGTGAAGCGCGCATGGAAGCG 726
Db	661	AATTTCCCATCAGCCCCCATCGAGACCGTGGCGTGAAGTGAAGCGCGCATGGAAGCG 720
Qy	727	CCCAGGTGAGCAGTGGCGCGCTGACCGAGGAGAGATCAAGGCGCTTACCGCATCTGC 786
Db	721	CCCAGGTGAGCAGTGGCGCGCTGACCGAGGAGAGATCAAGGCGCTTACCGCATCTGC 780
Qy	787	GAGGAGTGAAGAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCG 846
Db	781	GAGGAGTGAAGAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCG 840
Qy	847	CCCGTGTTCGCGATCAAGAGAGGAGCAGCAACAGTGGCGCAAGTGGTGGACTTCCG 906
Db	841	CCCGTGTTCGCGATCAAGAGAGGAGCAGCAACAGTGGCGCAAGTGGTGGACTTCCG 900
Qy	907	GAGTGAACAGCGCACCGCGAGCTTCTGGGAGGTGAGCTGGGCGATCCCCCGCGCG 966
Db	901	GAGTGAACAGCGCACCGCGAGCTTCTGGGAGGTGAGCTGGGCGATCCCCCGCGCG 960

Qy	967	GCCCTGAAGAAAGAAAGAGAGCGTGAACCTGTGGACGTGGCGAGCGCCTTACTTCAAGCTG 1026
Db	961	GCCCTGAAGAAAGAAAGAGAGCGTGAACCTGTGGACGTGGCGAGCGCCTTACTTCAAGCTG 1020
Qy	1027	CCCTTGAAGAGGAGTTCGCGCAAGTACACCGCTTACCATCCCGAGCATCAACACGAG 1086
Db	1021	CCCTTGAAGAGGAGTTCGCGCAAGTACACCGCTTACCATCCCGAGCATCAACACGAG 1080
Qy	1087	ACCCCGGCACTCCGCTACAGTACAACTGTGCTGCCCGAGGGCTGGAAGGGAGAGCGCCAGC 1146
Db	1081	ACCCCGGCACTCCGCTACAGTACAACTGTGCTGCCCGAGGGCTGGAAGGGAGAGCGCCAGC 1140
Qy	1147	ATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db	1141	ATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy	1207	GTGATCTTACAGGCGCCCTCTGTAGTGGGAGCGAGCCTGGAAGATCGGCGAGCAGCCGCG 1266
Db	1201	GTGATCTTACAGGCGCCCTCTGTAGTGGGAGCGAGCCTGGAAGATCGGCGAGCAGCCGCG 1260
Qy	1267	AAGATCGAGGAGCTGCGCAAGCAGCTGTGCGCTGGGGCTTTCACCAACCCCGCAAGAG 1326
Db	1261	AAGATCGAGGAGCTGCGCAAGCAGCTGTGCGCTGGGGCTTTCACCAACCCCGCAAGAG 1320
Qy	1327	CACAGAGAGGAGCGCCCTTCTGTGATGGGCTACGAGCTGCACCCCGCAAGTGGAGC 1386
Db	1321	CACAGAGAGGAGCGCCCTTCTGTGATGGGCTACGAGCTGCACCCCGCAAGTGGAGC 1374
Qy	1387	GTGAGCCCATCGAGCTGCGCGAGAGAGAGCTGGAGCGTGAACGATCCAGAGAGCTG 1446
Db	1375	GTGAGCCCATCGAGCTGCGCGAGAGAGAGCTGGAGCGTGAACGATCCAGAGAGCTG 1434
Qy	1447	GTGGGCAAGCTGAATCTGGGCGAGCAGATCTACCCCGGCGCATCAAGGTGCGCGAGCTG 1506
Db	1435	GTGGGCAAGCTGAATCTGGGCGAGCAGATCTACCCCGGCGCATCAAGGTGCGCGAGCTG 1494
Qy	1507	AAGTGTGTCGCGCGCGCAAGCGCTGACCAATCTGTGCGCCCTGACCGAGAGGAGCGCG 1566
Db	1495	AAGTGTGTCGCGCGCGCAAGCGCTGACCAATCTGTGCGCCCTGACCGAGAGGAGCGCG 1554
Qy	1567	CTGAGCTGCGCGAGAGCGCGAGATCTGCGCGAGCGCGTGCACCGGCTGTACTACGAC 1626
Db	1555	CTGAGCTGCGCGAGAGCGCGAGATCTGCGCGAGCGCGTGCACCGGCTGTACTACGAC 1614
Qy	1627	CCAGAGAGGAGCTGGTGGCGAGATCCAGAGAGAGGCGCACAGCAGTGAACCTTACGAG 1686
Db	1615	CCAGAGAGGAGCTGGTGGCGAGATCCAGAGAGAGGCGCACAGCAGTGAACCTTACGAG 1674
Qy	1687	ATCTACCAAGGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCG 1746
Db	1675	ATCTACCAAGGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCG 1734
Qy	1747	CACACCAAGAGCTGAAGCAGCTGACGAGCGCGTGCAGAGATCGCCATCGAGAGAGATC 1806
Db	1735	CACACCAAGAGCTGAAGCAGCTGACGAGCGCGTGCAGAGATCGCCATCGAGAGAGATC 1794
Qy	1807	GTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGAGAGCTGGGAGAG 1866
Db	1795	GTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGAGAGCTGGGAGAG 1854
Qy	1867	TGGTGGAGCGACTACTTGGCAGGCGCACCTGGATCCCCCGAGTGGGAGTTCGTGAACACCC 1926
Db	1855	TGGTGGAGCGACTACTTGGCAGGCGCACCTGGATCCCCCGAGTGGGAGTTCGTGAACACCC 1914
Qy	1927	CCCTGTGTGAAGCTGTGTATCAGCTGAGAGAGGAGCGCCATCATCGGCGCGAGAGCTTTC 1986
Db	1915	CCCTGTGTGAAGCTGTGTATCAGCTGAGAGAGGAGCGCCATCATCGGCGCGAGAGCTTTC 1974
Qy	1987	TACGTGAAGCGCGCGCGCAACCGCGAGAGAGAGTTCGGAAGGCGCGCTAGTGAACCGAG 2046
Db	1975	TACGTGAAGCGCGCGCGCAACCGCGAGAGAGAGTTCGGAAGGCGCGCTAGTGAACCGAG 2034



2047 CGGGCGCGCAGAGATCGTGAGCCTGACCGAGACCAACACAGAGACCGAGCTGCAG 2106  
 2035 CGGGCGCGCAGAGATCGTGAGCCTGACCGAGACCAACACAGAGACCGAGCTGCAG 2094  
 2107 GCCATCAGCTGGCCCTGCGAGGACAGCGGAGGAGGTGAACATCTGTACCGACAGCCAG 2166  
 2095 GCCATCAGCTGGCCCTGCGAGGACAGCGGAGGAGGTGAACATCTGTACCGACAGCCAG 2154  
 2167 TAGCCCTTGGGCATCTCAGGCGCCAGCCGACAGAGCGAGCGAGCTGTGAACCCAG 2226  
 2155 TAGCCCTTGGGCATCTCAGGCGCCAGCCGACAGAGCGAGCGAGCTGTGAACCCAG 2214  
 2227 ATCATCAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCCCAAG 2286  
 2215 ATCATCAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCCCAAG 2274  
 2287 GGCAATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGECATCCGCAAGTGTG 2346  
 2275 GGCAATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGECATCCGCAAGTGTG 2334  
 2347 TTCTTGACCGCATCGATGCGGCGATCGTATCTACAGTACATGGAAGCAGCTGTAGCTG 2406  
 2335 TTCTTGACCGCATCGATGCGGCGATCGTATCTACAGTACATGGAAGCAGCTGTAGCTG 2394  
 2407 GGCAGCGCGCGCTAGCATCGATTAAGGCTTCCCGGGCTAGCACCGGT 2457  
 2395 GGCAGCGCGCGCTAGCATCGATTAAGGCTTCCCGGGCTAGCACCGGT 2445

RESULT 13  
 ADCL3230  
 ID ADCL3230 standard; DNA; 3930 BP.  
 XX AC ADCL3230;  
 XX DT 18-DEC-2003 (first entry)  
 XX DE DNA of HIV construct GagCompIPolmut\_C SEQ ID NO 9.  
 XX KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX OS Human immunodeficiency virus.  
 XX PN W02003004620-A2.  
 XX PD 16-JAN-2003.  
 XX PF 05-JUL-2002; 2002WO-US021420.  
 XX PR 05-JUL-2001; 2001US-0303192P.  
 XX PR 31-AUG-2001; 2001US-0316860P.  
 XX PR 16-JAN-2002; 2002US-0349871P.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PA (UYST-) UNIV STELLENBOSCH.  
 XX ZUR Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 XX WPI; 2003-221593/21.  
 XX PT New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.  
 XX PS Disclosure; Fig 6; 301pp; English.  
 XX CC The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is

CC administered intramuscularly, intravenously, intranasally,  
 CC subcutaneously, intradermally, transdermally, intravaginally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.  
 XX SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;  
 Query Match 98.1%; Score 2415.6; DB 10; Length 3930;  
 Best Local Similarity 99.8%; Pred. No. 2.8e-291;  
 Matches 2434; Conservative 0; Mismatches 4; Indels 6; Gaps 1;  
 QY 14 TGGCCGAGGCGCATGAGCCAGGCCACACAGCGCCCAACATCTGTATGCGAGCGCAGCACTTCA 73  
 DB 1487 TCGCCGAGGCCATGAGCCAGGCCACACAGCGCCCAACATCTGTATGCGAGCGCAGCACTTCA 1546  
 QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGCA 133  
 DB 1547 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATCGCCCGCA 1606  
 QY 134 ACTGCGCGCGCCCCCGCAGAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACAGATGA 193  
 DB 1607 ACTGCGCGCGCCCCCGCAGAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACAGATGA 1666  
 QY 194 AGGACTGCAACCGAGCGCCAGGCCAACTTTCTTCCGCGAGGACCTTGGGCTTCCCCCAGGGCA 253  
 DB 1667 AGGACTGCAACCGAGCGCCAGGCCAACTTTCTTCCGCGAGGACCTTGGGCTTCCCCCAGGGCA 1726  
 QY 254 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCCAACAGCGCGAGTGC 313  
 DB 1727 AGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCCAACAGCGCGAGTGC 1786  
 QY 314 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCCGCGCGAGCGCGGCGAGGCGCAGCTGAAC 373  
 DB 1787 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCCGCGCGAGCGCGGCGAGGCGCAGCTGAAC 1846  
 QY 374 TCCCCCAGATCACCTGTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGCAGATCA 433  
 DB 1847 TCCCCCAGATCACCTGTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGCAGATCA 1906  
 QY 434 AGAGGCGCTGTGTGAGCACCGCGCGCGACGACACCGTGTGTGAGGAGATGAGCTGCGCG 493  
 DB 1907 AGAGGCGCTGTGTGAGCACCGCGCGCGACGACACCGTGTGTGAGGAGATGAGCTGCGCG 1966  
 QY 494 GCAAGTGGAGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAGGTGCGCGAGTACG 553  
 DB 1967 GCAAGTGGAGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAGGTGCGCGAGTACG 2026  
 QY 554 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCGACCGTGTGTGATGCGGCCCA 613  
 DB 2027 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCGACCGTGTGTGATGCGGCCCA 2086  
 QY 614 CCCCCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGACCCCTGAGCTTCC 673  
 DB 2087 CCCCCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGACCCCTGAGCTTCC 2146  
 QY 674 CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAGCGCGCATGAGCGCGCGCATGAGCGGCCAAGG 733  
 DB 2147 CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAGCGCGCATGAGCGCGCGCATGAGCGGCCAAGG 2206  
 QY 734 TGAAGCAGTGGCCCTTGACCGAGGAGAAAGATCAAGGCCCTTGAGCCGCATCTGCGAGGAGA 793  
 DB 2207 TGAAGCAGTGGCCCTTGACCGAGGAGAAAGATCAAGGCCCTTGAGCCGCATCTGCGAGGAGA 2266  
 QY 794 TGGAGAGGAGGGCAAGATCAACCAAGATCGCGCGCGAGAACCCCTTACAAACCCCGCTGT 853  
 DB 2267 TGGAGAGGAGGGCAAGATCAACCAAGATCGCGCGCGAGAACCCCTTACAAACCCCGCTGT 2326  
 QY 854 TCGCCATCAGAGAGAGGACAGCAACCAAGTGGCGCAGCTGTGAGCTTCCGGGAGCTGA 913  
 DB 2327 TCGCCATCAGAGAGAGGACAGCAACCAAGTGGCGCAGCTGTGAGCTTCCGGGAGCTGA 2386

QY 914 ACAAGCCGACCCAGAGACTTCTGGGAGGTGAGCTGGGATCCCCACACCCCGCGGCTGA 973  
 Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGATCCCCACACCCCGCGGCTGA 2446  
 QY 974 AGAAGAACAGAGCTGACCGTCTGGACGTGGGCGAGCCCTACTTCAAGCTGGCCCTGG 1033  
 Db 2447 AGAAGAACAGAGCTGACCGTCTGGACGTGGGCGAGCCCTACTTCAAGCTGGCCCTGG 2506  
 QY 1034 ACAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAACAGAGACCCCG 1093  
 Db 2507 ACAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAACAGAGACCCCG 2566  
 QY 1094 GCATCCGCTACAGTACAAAGTGTGCCCCAGGGCTGGAAGGCGAGCCCGAGCATCTTCC 1153  
 Db 2567 GCATCCGCTACAGTACAAAGTGTGCCCCAGGGCTGGAAGGCGAGCCCGAGCATCTTCC 2626  
 QY 1154 AGAGCAGCATCAACAAAGTCTTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGATCT 1213  
 Db 2627 AGAGCAGCATCAACAAAGTCTTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGATCT 2686  
 QY 1214 ACCAGGCCCCCTGTACGTGGGCGAGGACCTTGGAGATCGGCAGCACCGCGCCCAAGATCG 1273  
 Db 2687 ACCAGGCCCCCTGTACGTGGGCGAGGACCTTGGAGATCGGCAGCACCGCGCCCAAGATCG 2746  
 QY 1274 AGGAGCTGCGCAAGCACTGTGCGCTGGGGTTTCAACCCCGCAAGAGCAACAGA 1333  
 Db 2747 AGGAGCTGCGCAAGCACTGTGCGCTGGGGTTTCAACCCCGCAAGAGCAACAGA 2806  
 QY 1334 AGGAGCCCCCTTCTGTGGATGGGTACGAGCTGCACCCCGCAAGTGGACCGTGCAGC 1393  
 Db 2807 AGGAGCCCCCTTCTGTGGATGGGTACGAGCTGCACCCCGCAAGTGGACCGTGCAGC 2860  
 QY 1394 CCATCGAGCTGCCGAGAGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGGTGGCA 1453  
 Db 2861 CCATCGAGCTGCCGAGAGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGGTGGCA 2920  
 QY 1454 AGCTGAACCTGGGCGAGCCAGATCTACCCCGGATCAAGGTGGCGAGCTGTGCAAGTGC 1513  
 Db 2921 AGCTGAACCTGGGCGAGCCAGATCTACCCCGGATCAAGGTGGCGAGCTGTGCAAGTGC 2980  
 QY 1514 TGCGGGGCGCAAGGCGCTGACGACATCGTCCCTGACCGAGGAGGCGCGAGCTGGAGC 1573  
 Db 2981 TGCGGGGCGCAAGGCGCTGACGACATCGTCCCTGACCGAGGAGGCGCGAGCTGGAGC 3040  
 QY 1574 TGCGCGGAACCCGCGAGATCTTGGCGGAGCCGCTGACGCGGTGTACTACGACCCCAAGCA 1633  
 Db 3041 TGCGCGGAACCCGCGAGATCTTGGCGGAGCCGCTGACGCGGTGTACTACGACCCCAAGCA 3100  
 QY 1634 AGGACCTGGTGGCGAGATCCAGAAAGCGAGGCGCCAGCACTGGACCTTACATCTACC 1693  
 Db 3101 AGGACCTGGTGGCGAGATCCAGAAAGCGAGGCGCCAGCACTGGACCTTACATCTACC 3160  
 QY 1694 AGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCAAGATCGGACCGCCGACACCA 1753  
 Db 3161 AGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCAAGATCGGACCGCCGACACCA 3220  
 QY 1754 ACAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCCATCGAGAGCATCTGTATCT 1813  
 Db 3221 ACAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCCATCGAGAGCATCTGTATCT 3280  
 QY 1814 GGGGCAAGACCCCAAGTTCGCTGCGCCATCCAGAAAGGAGACTGGGAGACCTGGTGA 1873  
 Db 3281 GGGGCAAGACCCCAAGTTCGCTGCGCCATCCAGAAAGGAGACTGGGAGACCTGGTGA 3340  
 QY 1874 CCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTGTTGAACACCCCGCCCTGG 1933  
 Db 3341 CCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTGTTGAACACCCCGCCCTGG 3400  
 QY 1934 TGAAGCTGTGTATACAGCTGAGAGGAGCCCATCATCTCGGCGCGAGACCTTCTACGTGG 1993  
 Db 3401 TGAAGCTGTGTATACAGCTGAGAGGAGCCCATCATCTCGGCGCGAGACCTTCTACGTGG 3460  
 QY 1994 ACGGCGCGCCACCGGAGACCAAGATCGGCAAGCGCGGCTACGTGACCGCGGGGCC 2053

Db 3461 ACGGCGCGCCACCGGAGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGGGCC 3520  
 QY 2054 GGCAGAGATCTGAGCCCTGAGCCGAGACCAACCAAGAACCCAGCTGCAAGGCGCATCC 2113  
 Db 3521 GGCAGAGATCTGAGCCCTGAGCCGAGACCAACCAAGAACCCAGCTGCAAGGCGCATCC 3580  
 QY 2114 AGCTGGCCCTCAGGACAGCGGACGAGGTTGAACATCTGTAACGACAGGATACGCC 2173  
 Db 3581 AGCTGGCCCTCAGGACAGCGGACGAGGTTGAACATCTGTAACGAGCATCATCG 3640  
 QY 2174 TGGGATCATCCAGGCGCCAGCCGACCAAGAGGAGGAGCTGGTGAACAGATCATCG 2233  
 Db 3641 TGGGATCATCCAGGCGCCAGCCGACCAAGAGGAGGAGCTGGTGAACAGATCATCG 3700  
 QY 2234 AGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCCAAGGGCATCG 2293  
 Db 3701 AGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCCAAGGGCATCG 3760  
 QY 2294 GCGCAACGAGCAGATCGACAAAGCTGGTGAAGAGGCGATCCGCAAGGTGCTGTTCTGG 2353  
 Db 3761 GCGCAACGAGCAGATCGACAAAGCTGGTGAAGAGGCGATCCGCAAGGTGCTGTTCTGG 3820  
 QY 2354 ACGGATCATGGCGGCGATCTGATCTACAGTACATGAGCAGCTGTACGTGGGCGAGC 2413  
 Db 3821 ACGGATCATGGCGGCGATCTGATCTACAGTACATGAGCAGCTGTACGTGGGCGAGC 3880  
 QY 2414 GCGGCCCTAGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 2457  
 Db 3881 GCGGCCCTAGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 14  
 ADC13231  
 ID ADC13231 standard; DNA; 3930 BP.  
 XX AC ADC13231;  
 XX DT 18-DEC-2003 (first entry)  
 XX DE DNA of HIV construct GagCompIPolmutAtt\_C SEQ ID NO 10.  
 XX KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 XX KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX OS Human immunodeficiency virus.  
 XX FN WO2003004620-A2.  
 XX PD 16-JAN-2003.  
 XX PF 05-JUL-2002; 2002WO-US021420.  
 XX PR 05-JUL-2001; 2001US-0303192P.  
 XX PR 31-AUG-2001; 2001US-0316860P.  
 XX PR 16-JAN-2002; 2002US-0349871P.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PA (UYST-) UNIV STELLENBOSCH.  
 XX PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ;  
 XX DR WPI; 2003-221593/21.  
 XX PT New expression cassette comprising a polynucleotide sequence encoding a  
 XX PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 XX PT Prot, or Rev polypeptide, useful for immunization, or generating  
 XX PT packaging cell lines.  
 XX PS Disclosure; Fig 7; 301pp; English.  
 XX CC The invention relates to a novel expression cassette comprising a  
 XX CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,

CC	Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel	
CC	expression cassette can be used to treat HIV type C by gene therapy or	
CC	used in the development of a vaccine. The gene delivery vector is	
CC	administered intramuscularly, intramuscularly, intranasally,	
CC	subcutaneously, intradermally, transdermally, intravaginally,	
CC	intrarectally, orally or intravenously. The expression cassette is useful	
CC	for immunisation, generating packaging cell lines and producing HIV	
CC	polypeptides. This polynucleotide sequence represents the DNA of an HIV	
CC	Type C related sequence of the invention.	
XX		
QQ	Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;	
	Query Match 98.0%; Score 2414; DB 10; Length 3930;	
	Best Local Similarity 99.5%; Pred. No. 4.4e-291;	
	Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;	
QY	14 TGGCCGAGGCGCATGAGCCAGGCGCACAGCGCCCAACATCTTGATGCGAGCGCAACTTCA 73	
DB	1487 TCGCCGAGGCGCATGAGCCAGGCGCACAGCGCCCAACATCTTGATGCGAGCGCAACTTCA 1546	
QY	74 AGGCCCCCAAGCGCATCATCAAGTGTCTCACTGCGCGAAGGAGGCGCATGCGCCGCA 133	
DB	1547 AGGCCCCCAAGCGCATCATCAAGTGTCTCACTGCGCGAAGGAGGCGCATGCGCCGCA 1606	
QY	134 ACTGCCGCGCCCCCGCAAGAAGGGCTGTGGAGTGGCGCAAGGAGGCGCACAGATGA 193	
DB	1607 ACTGCCGCGCCCCCGCAAGAGGGCTGTGGAGTGGCGCAAGGAGGCGCACAGATGA 1666	
QY	194 AGGACTGCAACGAGCGCGAGGCCAACTTCTTTCGCGAGGACCTTGGCTTCCCCAGGGCA 253	
DB	1667 AGGACTGCAACGAGCGCGAGGCCAACTTCTTTCGCGAGGACCTTGGCTTCCCCAGGGCA 1726	
QY	254 AGGCCCGGAGTTCCTCAGCGAGCGAGAACCGCGCCACAGCCGCCACAGCGCGAGCTGC 313	
DB	1727 AGGCCCGGAGTTCCTCAGCGAGCGAGAACCGCGCCACAGCCGCCACAGCGCGAGCTGC 1786	
QY	314 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGAGCGCACCTGAACT 373	
DB	1787 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGAGCGCACCTGAACT 1846	
QY	374 TCCCCAGATACCTCTGGAGCGCCCCCTGTGTGAGCATCAAGTGGGCGCGCGAGATCA 433	
DB	1847 TCCCCAGATACCTCTGGAGCGCCCCCTGTGTGAGCATCAAGTGGGCGCGCGAGATCA 1906	
QY	434 AGGAGCCCTGTCTGGAACAACCGCGCGCGAGCACACCGTGTGGAGGATGAGCTGCGCG 493	
DB	1907 AGGAGCCCTGTCTGGAACAACCGCGCGCGAGCACACCGTGTGGAGGATGAGCTGCGCG 1966	
QY	494 GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGAGTACG 553	
DB	1967 GCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGAGTACG 2026	
QY	554 ACCAGATCTCTGATCGAGATCTGGCGCAAGAGCGCATCGGCACCGTGTGATCGGCCCA 613	
DB	2027 ACCAGATCTCTGATCGAGATCTGGCGCAAGAGCGCATCGGCACCGTGTGATCGGCCCA 2086	
QY	614 CCCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGACCCCTGNACTTCC 673	
DB	2087 CCCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGACCCCTGNACTTCC 2146	
QY	674 CCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGGAAGCGCGCATCGAGCGGCCCAAGG 733	
DB	2147 CCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGGAAGCGCGCATCGAGCGGCCCAAGG 2206	
QY	734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGA 793	
DB	2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGA 2266	
QY	794 TGGAGAGGAGGCGAAGATCACCAAGATCGGCCCGCGAGACCCCTTACAACCCCGCTGT 853	
DB	2267 TGGAGAGGAGGCGGCAAGATCACCAAGATCGGCCCGCGAGACCCCTTACAACCCCGCTGT 2326	
QY	854 TCGCCATCAAGAGAGGAGCAGCAACCAAGTGGCGCGAAGCTGGTGGACTTCCGCGAGCTGA 913	

DB	2327 TCGCCATCAAGAGAAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGA 2386	
QY	914 ACAAGCGCACCCAGGACTTCTGGAGGTGCAGTGGGCATCCCCACCCCGCGGCTGA 973	
DB	2387 ACAAGCGCACCCAGGACTTCTGGAGGTGCAGTGGGCATCCCCACCCCGCGGCTGA 2446	
QY	974 AGAAGAGAGAGCGTGCAGCGTCTGGAAGTGGGCGACGCTACTTTCAGCGTGGCCCTGG 1033	
DB	2447 AGAAGAGAGAGCGTGCAGCGTCTGGAAGTGGGCGACGCTACTTTCAGCGTGGCCCTGG 2506	
QY	1034 ACAGGACTTTCGCAAGTACACCGCTTCAACATCCCCAGCATCAACAAACAGACCCCG 1093	
DB	2507 ACAGGACTTTCGCAAGTACACCGCTTCAACATCCCCAGCATCAACAAACAGACCCCG 2566	
QY	1094 GCATCGCTACAGTACACAGCTGCTGCCCGCAGGGCTGGAGGGCAGCCCGAGCATCTTCC 1153	
DB	2567 GCATCGCTACAGTACACAGCTGCTGCCCGCAGGGCTGGAGGGCAGCCCGAGCATCTTCC 2626	
QY	1154 AGAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCGCAACCCCGAGATCGTATCT 1213	
DB	2627 AGAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCGCAACCCCGAGATCGTATCT 2686	
QY	1214 ACCAGGCCCTCTGTACGTGGGAGCAGCACTTGAGATTCGGCCAGCACCGCGCAAGATCG 1273	
DB	2687 ACCAGGCCCTCTGTACGTGGGAGCAGCACTTGAGATTCGGCCAGCACCGCGCAAGATCG 2746	
QY	1274 AGAGCTGGCAAGCAGCTGCTGGTGGGGCTTCAACCCCGCAACAGAACCCAGCA 1333	
DB	2747 AGAGCTGGCAAGCAGCTGCTGGTGGGGCTTCAACCCCGCAACAGAACCCAGCA 2806	
QY	1334 AGGAGCCCCCTTCTGTGGATGGGTACAGAGCTGCACCCCGCAACAGTGAACCGTGAGC 1393	
DB	2807 AGGAGCCCCCTTCTGTGGATGGGTACAGAGCTGCACCCCGCAACAGTGAACCGTGAGC 2860	
QY	1394 CCATCGAGTGTCCCGAAGAGGAGCTGAGACCTGTGAACCGACATCCAGAGCTGGTGGGCA 1453	
DB	2861 CCATCGAGTGTCCCGAAGAGGAGCTGAGACCTGTGAACCGACATCCAGAGCTGGTGGGCA 2920	
QY	1454 AGCTGAATGGGCGACCGCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGAAGCTGC 1513	
DB	2921 AGCTGAATGGGCGACCGCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGAAGCTGC 2980	
QY	1514 TCGCGCGCGCAAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAGCGCGAGCTGAGC 1573	
DB	2981 TCGCGCGCGCAAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAGCGCGAGCTGAGC 3040	
QY	1574 TGGCCGAGAACCCCGAGATCTCTGCGAGCGCCCGTGCACCGCGTGTACTACGACCCAGCA 1633	
DB	3041 TGGCCGAGAACCCCGAGATCTCTGCGAGCGCCCGTGCACCGCGTGTACTACGACCCAGCA 3100	
QY	1634 AGGACTGGTGGCGGAGATCCAGAGCGGCGCACGACCAAGTGGAGCTTACAGATCTTACC 1693	
DB	3101 AGGACTGGTGGCGGAGATCCAGAGCGGCGCACGACCAAGTGGAGCTTACAGATCTTACC 3160	
QY	1694 AGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCAAGATGCGCACCGCCACACCA 1753	
DB	3161 AGGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCAAGATGCGCACCGCCACACCA 3220	
QY	1754 ACAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATTCGCATCGAGAGAGCATCGTATCT 1813	
DB	3221 ACAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATTCGCATCGAGAGAGCATCGTATCT 3280	
QY	1814 GGGCAGAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGAGCTGGGAGACCTTGGTGGGA 1873	
DB	3281 GGGCAGAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGAGCTGGGAGACCTTGGTGGGA 3340	
QY	1874 CCAGTACTGGCAGGCGCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGG 1933	
DB	3341 CCAGTACTGGCAGGCGCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGG 3400	
QY	1934 TGAAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGCTGG 1993	

Db 3401 TGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGGCCGAGACCTTCTACGTGG 3460

Qy 1994 ACGGCGCCGCAACCGCAGACCAAGATCGGAGCGCGGTACTGTCAGCCGAGCGGCC 2053

Db 3461 ACGGCGCCGCAACCGCAGACCAAGATCGGAGCGCGGTACTGTCAGCCGAGCGGCC 3520

Qy 2054 GGCAGAAATGCTGAGCTTACCGGAGACCAACCAAGAGACCGAGTGCAGGCGCATCC 2113

Db 3521 GGCAGAAATGCTGAGCTTACCGGAGACCAACCAAGAGACCGAGTGCAGGCGCATCC 3580

Qy 2114 AGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTACCGACAGCCAGTACGCCC 2173

Db 3581 AGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTACCGACAGCCAGTACGCCC 3640

Qy 2174 TGGGCATCATCCAGGCCCGCCGCAAGAGCGAGCGAGTGGTGAACAGATCATCG 2233

Db 3641 TGGGCATCATCCAGGCCCGCCGCAAGAGCGAGCGAGTGGTGAACAGATCATCG 3700

Qy 2234 AGCAGCTGATCAAGAGAGAGAGGTGTACTGTAGCTGGGTGGCCGCCACAAAGGCGATCG 2293

Db 3701 AGCAGCTGATCAAGAGAGAGAGGTGTACTGTAGCTGGGTGGCCGCCACAAAGGCGATCG 3760

Qy 2294 GCGGCAAGCAGCAGATCCACAAGCTGGTGAGCAAGGCGATCCGCAAGGTGCTGTTCCTGG 2353

Db 3761 GCGGCAAGCAGCAGATCCACAAGCTGGTGAGCAAGGCGATCCGCAAGGTGCTGTTCCTGG 3820

Qy 2354 ACGGCATCGATGGCGGCGATCGTGATCTACCAAGTACATGGAGCGACCTGTACGTGGCGAGCG 2413

Db 3821 ACGGCATCGATGGCGGCGATCGTGATCTACCAAGTACATGGAGCGACCTGTACGTGGCGAGCG 3880

Qy 2414 GCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457

Db 3881 GCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 15

ADCL13232

ID ADCL13232 standard; DNA; 3930 BP.

XX AC ADCL13232;

XX DT 18-DEC-2003 (first entry)

XX DE DNA of HIV construct GagComplPolmutIna\_C SEQ ID NO 11.

XX KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;

XX KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX OS Human immunodeficiency virus.

XX PN WO2003004620-A2.

XX PD 16-JAN-2003.

XX PF 05-JUL-2002; 2002WO-US021420.

XX PR 05-JUL-2001; 2001US-0303192P.

XX PR 31-AUG-2001; 2001US-0316860P.

XX PR 16-JAN-2002; 2002US-0349871P.

XX PA (CHIR ) CHIRON CORP.

XX PA (UYST-) UNIV STELLENBOSCH.

XX PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX DR WPI; 2003-221593/21.

XX PT New expression cassette comprising a polynucleotide sequence encoding a

XX PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

XX PT Prot, or Rev polypeptide, useful for immunization, or generating

XX PT packaging cell lines.

XX PS Disclosure; Fig 8; 301pp; English.

XX The invention relates to a novel expression cassette comprising a

CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,

CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel

CC expression cassette can be used to treat HIV type C by gene therapy or

CC used in the development of a vaccine. The gene delivery vector is

CC administered intramuscularly, intranasally, intravenously,

CC subcutaneously, intradermally, transdermally, or intravenously. The

CC for immunisation, generating packaging cell lines and producing HIV

CC polypeptides. This polynucleotide sequence represents the DNA of an HIV

CC Type C related sequence of the invention.

XX SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 98.0%; Score 2414; DB 10; Length 3930;

Best Local Similarity 99.5%; Pred. No. 4.4e-291;

Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

Qy 14 TGGCGAGGCGCATGAGCGGCGCCACGCGCAACATCTGTATGAGCGCGCAACTTCA 73

Db 1487 TCGCGAGGCGCATGAGCGGCGCCACGCGCAACATCTGTATGAGCGCGCAACTTCA 1546

Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGGCGAAGGCGGCACATCGCCCGCA 133

Db 1547 AGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGGCGAAGGCGGCACATCGCCCGCA 1606

Qy 134 ACTGCGCGCGCCCCCGCAAGAGGCGTCTCGAAGTGGCGAAGGCGGCACACAGATGA 193

Db 1607 ACTGCGCGCGCCCCCGCAAGAGGCGTCTCGAAGTGGCGAAGGCGGCACACAGATGA 1666

Qy 194 AGGACTGCAACGAGCGCGCGCCCAACTTTCTTCGCGAGGACCTGGCCCTTCCCGCAAGGCA 253

Db 1667 AGGACTGCAACGAGCGCGCGCCCAACTTTCTTCGCGAGGACCTGGCCCTTCCCGCAAGGCA 1726

Qy 254 AGGCGCGCGAGTTCGCGAGCGAGCAGACCGCGCCCAAGCGGCACACAGCGCGCGAGTGC 313

Db 1727 AGGCGCGCGAGTTCGCGAGCGAGCAGACCGCGCCCAAGCGGCACACAGCGCGCGAGTGC 1786

Qy 314 AGGTGCGCGCGCAACACCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373

Db 1787 AGGTGCGCGCGCAACACCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846

Qy 374 TCCCCCAGATCACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433

Db 1847 TCCCCCAGATCACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906

Qy 434 AGGAGGCGCGTCTGCGACCG 493

Db 1907 AGGAGGCGCGTCTGCGACCG 1966

Qy 494 GCAAGTGGAGCGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGCGCGCGCG 553

Db 1967 GCAAGTGGAGCGCCCAAGATGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGCGCGCG 2026

Qy 554 ACCAGATCTCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGACCGTGTGATCGCGCGCG 613

Db 2027 ACCAGATCTCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGACCGTGTGATCGCGCGCG 2086

Qy 614 CCCCCGTGAAACATCATCGCGCGCGCAACATGCTGACCCAGCTGGGCTGCGACCTGAACTTCC 673

Db 2087 CCCCCGTGAAACATCATCGCGCGCGCAACATGCTGACCCAGCTGGGCTGCGACCTGAACTTCC 2146

Qy 674 CCATCAGGCGCGCATCGAGACCGTGGCGCGTGAAGCGCGCGCATGGAAGCGCGCGCGCGCG 733

Db 2147 CCATCAGGCGCGCATCGAGACCGTGGCGCGTGAAGCGCGCGCATGGAAGCGCGCGCGCG 2206

Qy 734 TGAAGCAGTGGCG 793

Db 2207 TGAAGCAGTGGCG 2266

Qy 794 TGGAGAGGAGGCG 853

Db	2267	TGAGAAAGGAGGCAAGATCAACAAGATCGCGCCCGGAGAACCCCTACAAACACCCCGTGT	2326
Qy	854	TCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAGCTGCTGGAGCTTCGCGAGCTGA	913
Db	2327	TCGCGCATCAAGAGAGGAGCAGCACCAAGTGGCGCAGCTGCTGGAGCTTCGCGAGCTGA	2386
Qy	914	ACAAGCCGACCCAGGACTTCTGGAGAGTGCAGCTGGGCATCCCCACCCCGCGGCGCTGA	973
Db	2387	ACAAGCCGACCCAGGACTTCTGGAGAGTGCAGCTGGGCATCCCCACCCCGCGGCGCTGA	2446
Qy	974	AGAAGAAAGAGAGCTGACCGTGTGCACTGGGCGCAGCCTACTTTCAGCGTGCCTGTG	1033
Db	2447	AGAAGAAAGAGAGCTGACCGTGTGCACTGGGCGCAGCCTACTTTCAGCGTGCCTGTG	2506
Qy	1034	ACGAGGACTTTCGCAAGTACACCGCTTTCACCTCCCGCAGCATCAACAAAGAGACCCCG	1093
Db	2507	ACGAGGACTTTCGCAAGTACACCGCTTTCACCTCCCGCAGCATCAACAAAGAGACCCCG	2566
Qy	1094	GCATCCGCTACCAAGTACAAAGTGTGCTGCCCGCAGGCTGGAAGGGCAGCCCCAGCATCTTCC	1153
Db	2567	GCATCCGCTACCAAGTACAAAGTGTGCTGCCCGCAGGCTGGAAGGGCAGCCCCAGCATCTTCC	2626
Qy	1154	AGAGCAGCATGA CCAAGATCTTGAGGCTTTCGCGCCCGCAACCCCGAGATCGTGATCT	1213
Db	2627	AGAGCAGCATGA CCAAGATCTTGAGGCTTTCGCGCCCGCAACCCCGAGATCGTGATCT	2686
Qy	1214	ACNAGGCCCCCTGTAGTGGGCGAGCAGCTGGAGATCGGCGCAGCACCGGCCCAAGATCG	1273
Db	2687	ACNAGGCCCCCTGTAGTGGGCGAGCAGCTGGAGATCGGCGCAGCACCGGCCCAAGATCG	2746
Qy	1274	AGAGCTGCGCAAGCAGCTGCTGGCTGGGGCTTTCACCAACCCCGCAGCAAGAACCCAGCA	1333
Db	2747	AGAGCTGCGCAAGCAGCTGCTGGCTGGGGCTTTCACCAACCCCGCAGCAAGAACCCAGCA	2806
Qy	1334	AGGAGCCCCCTTCTCTGTGATGGGCTTACGAGCTGCAACCCCGCAAGTGGACCGTGACG	1393
Db	2807	AGGAGCCCCCTTCTCTGTGATGGGCTTACGAGCTGCAACCCCGCAAGTGGACCGTGACG	2860
Qy	1394	CCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAACTTGTGGGCA	1453
Db	2861	CCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAACTTGTGGGCA	2920
Qy	1454	AGCTGAATCGGGCCAGCCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCAAGCTGC	1513
Db	2921	AGCTGAATCGGGCCAGCCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCAAGCTGC	2980
Qy	1514	TGCGCGCGCCAAAGCCCTGACCGACATCGTGGCCCTTGA CCGAGAGGCGGAGCTGGAGC	1573
Db	2981	TGCGCGCGCCAAAGCCCTGACCGACATCGTGGCCCTTGA CCGAGAGGCGGAGCTGGAGC	3040
Qy	1574	TGGCCGAGAACCGCGAGATCTGCGCGAGCCCGTGGACACGGCGTGTACTAGACCCGAGCA	1633
Db	3041	TGGCCGAGAACCGCGAGATCTGCGCGAGCCCGTGGACACGGCGTGTACTAGACCCGAGCA	3100
Qy	1634	AGGACCTGGTGGCGGAGATCCAGAAAGCAGGCGCCACGACAGTGGACCTTACAGATCTACC	1693
Db	3101	AGGACCTGGTGGCGGAGATCCAGAAAGCAGGCGCCACGACAGTGGACCTTACAGATCTACC	3160
Qy	1694	AGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCAACCA	1753
Db	3161	AGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCAACCA	3220
Qy	1754	ACGACGTGAACGAGCTGACCGAGGCGCTGCAGAAAGATCGGCATGGAGAGCATCGTGATCT	1813
Db	3221	ACGACGTGAACGAGCTGACCGAGGCGCTGCAGAAAGATCGGCATGGAGAGCATCGTGATCT	3280
Qy	1814	GGGCAAGACCCCAAGTTCGCGCTGCCCTTCCAGAAAGGAGACCTGGGAGACCTTGGTGA	1873
Db	3281	GGGCAAGACCCCAAGTTCGCGCTGCCCTTCCAGAAAGGAGACCTGGGAGACCTTGGTGA	3340
Qy	1874	CCGACTACTGGCAGGCGCACTGGATCCCGAGTGGGAGTTCTGTGAAACCCCGCCCTGTG	1933
Db	3341	CCGACTACTGGCAGGCGCACTGGATCCCGAGTGGGAGTTCTGTGAAACCCCGCCCTGTG	3400

Search completed: June 1, 2005, 11:33:26  
Job time : 1275.33 secs

Qy	1934	TGAAGCTGTGTTACCACTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACCTGG	1993
Db	3401	TGAAGCTGTGTTACCACTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACCTGG	3460
Qy	1994	ACGGCGCCGCAACCCGCGAGACCAAGATCGGCAAGGCGGCTACTGTGA CCGAGCGGGCC	2053
Db	3461	ACGGCGCCGCAACCCGCGAGACCAAGATCGGCAAGGCGGCTACTGTGA CCGAGCGGGCC	3520
Qy	2054	GGCAGAAAGATCGTGAGGCTGACCGAGACCAACCAACAGAAAGACCGAGCTCGAGGCCATCC	2113
Db	3521	GGCAGAAAGATCGTGAGGCTGACCGAGACCAACCAACAGAAAGACCGAGCTCGAGGCCATCC	3580
Qy	2114	AGCTGGCCCTGCAAGGACAGCGGCGAGGCTGAACATCGTGAC CCGAGCAGCCAGTACGCC	2173
Db	3581	AGCTGGCCCTGCAAGGACAGCGGCGAGGCTGAACATCGTGAC CCGAGCAGCCAGTACGCC	3640
Qy	2174	TGGGCATCATTCAGGGCCAGCCCGACAAGAGCGAGAGCTGGTGAACCAAGGGCATCG	2233
Db	3641	TGGGCATCATTCAGGGCCAGCCCGACAAGAGCGAGAGCTGGTGAACCAAGGGCATCG	3700
Qy	2234	AGCAGCTGATCAAGAAAGGAGAGGTTGTA CCGTGAGCTGGGTGCCGCCCAAGGGCATCG	2293
Db	3701	AGCAGCTGATCAAGAAAGGAGAGGTTGTA CCGTGAGCTGGGTGCCGCCCAAGGGCATCG	3760
Qy	2294	GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCCGCAAGGTGCTGTCCTGG	2353
Db	3761	GCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCCGCAAGGTGCTGTCCTGG	3820
Qy	2354	ACGGCATCGATGGCGGCATCGTGATCTTACAGTACATGGACGACCTGTGTACGTGGGACAGC	2413
Db	3821	ACGGCATCGATGGCGGCATCGTGATCTTACAGTACATGGACGACCTGTGTACGTGGGACAGC	3880
Qy	2414	GCGGCCCTTAGGATCGATTAAAAAGCTTCCCGGGGCTAGCACCGGT 2457	
Db	3881	GCGGCCCTTAGGATCGATTAAAAAGCTTCCCGGGGCTAGCACCGGT 3924	

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 03:50:55 ; Search time 10538.7 Seconds  
(without alignments)  
11324.506 Million cell updates/sec

Title: US-09-610-313B-31  
Perfect score: 2463  
Sequence: 1 gtcgacgcaccatggccga.....gggtagcaccggtagattc 2463

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_ats.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2463	100.0	2463	6	AX455915 Sequence
2	2442.2	99.2	2469	6	AX455914 Sequence
3	2436.2	98.9	2457	6	AX455916 Sequence
4	2046	83.1	2306	6	BD263704 Improved
5	2046	83.1	2306	6	CQ870574 Sequence
6	2046	83.1	2306	6	AR373387 Sequence
7	2027.6	82.3	9166	6	AX427930 Sequence
8	2025.2	82.2	2312	6	BD263706 Improved
9	2025.2	82.2	2312	6	CQ870576 Sequence
10	2025.2	82.2	2312	6	AR373389 Sequence
11	2019.2	82.0	2300	6	BD263705 Improved
12	2019.2	82.0	2300	6	CQ870575 Sequence
13	2019.2	82.0	2300	6	AR373388 Sequence
14	2019.2	82.0	9788	6	AX427936 Sequence
15	2005.2	81.4	9169	6	AX427931 Sequence
16	2001.8	81.3	9194	6	AX427926 Sequence
17	2000.2	81.2	9194	6	AX427925 Sequence
18	2000.2	81.2	12411	6	AX427927 Sequence
19	1996.8	81.1	9785	6	AX427938 Sequence

20	1993.4	80.9	9167	6	AX427933	AX427933 Sequence
21	1993.4	80.9	9170	6	AX427928	AX427928 Sequence
22	1993.4	80.9	9782	6	AX427935	AX427935 Sequence
23	1993.4	80.9	9783	6	AX427934	AX427934 Sequence
24	1993.4	80.9	9792	6	AX427932	AX427932 Sequence
25	1991.8	80.9	9189	6	AX427921	AX427921 Sequence
26	1988.8	80.7	4353	12	AF287353	AF287353 Synthetic
27	1972.2	80.1	3015	6	AX455946	AX455946 Sequence
28	1955.6	79.4	3009	6	AX455987	AX455987 Sequence
29	1942.6	78.9	4319	6	BD263639	BD263639 Improved
30	1942.6	78.9	4319	6	CQ870498	CQ870498 Sequence
31	1942.6	78.9	4319	6	AR373322	AR373322 Sequence
32	1932	78.4	4352	12	AF287352	AF287352 Synthetic
33	1886.6	76.6	2577	6	AX457088	AX457088 Sequence
34	1879	76.3	4343	6	AX306428	AX306428 Sequence
35	1878.8	76.3	2305	6	BD263702	BD263702 Improved
36	1878.8	76.3	2305	6	CQ870572	CQ870572 Sequence
37	1878.8	76.3	2305	6	AR373385	AR373385 Sequence
38	1875	76.1	4341	6	AX306429	AX306429 Sequence
39	1865.4	75.7	1977	6	AX455954	AX455954 Sequence
40	1857.8	75.4	1965	6	AX455952	AX455952 Sequence
41	1852.4	75.2	7897	6	AX427923	AX427923 Sequence
42	1852	75.2	2299	6	BD263703	BD263703 Improved
43	1852	75.2	2299	6	CQ870573	CQ870573 Sequence
44	1852	75.2	2299	6	AR373386	AR373386 Sequence
45	1798.8	73.0	3024	12	AF287355	AF287355 Synthetic

ALIGNMENTS

RESULT 1  
AX455915  
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 31 from Patent WO200493.  
ACCESSION AX455915  
VERSION AX455915.1 GI:21714900  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;  
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)  
FEATURES  
source  
1..2463  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YM"

ORIGIN

Query Match 100.0%; Score 2463; DB 6; Length 2463;  
Best Local Similarity 100.0%; Pred. No. 3.6e-248;  
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGACGCCACATGCGCGGCGCCATGAGCCAGGCCAACATCTCTGATGCAG 60  
DB 1 GTCGACGCCACATGCGCGGCGCCATGAGCCAGGCCAACATCTCTGATGCAG 60  
QY 61 CGCGCACTTTCAGGCGCCCAAGCGCATCATCACTGCTTCACTCGGCGGAGGCGC 120  
DB 61 CGCGCACTTTCAGGCGCCCAAGCGCATCATCACTGCTTCACTCGGCGGAGGCGC 120  
QY 121 CACATCGCGCCCAACTGCGCGCGCCCGCCGCAAGAGGCGTGTGGAAGTCGGCGCAAGGAG 180  
DB 121 CACATCGCGCCCAACTGCGCGCGCCCGCCGCAAGAGGCGTGTGGAAGTCGGCGCAAGGAG 180  
QY 181 GGCCACCAAGTGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGGAGGACTGGCC 240



Db 181 GGCACACAGATGAAGAGCTGCAACGAGCGCCAGGCCAACTTCTTCCGCGAGGAACTTG3CC 240  
Qy 241 TTCCCCAGAGCCAGGCGCGAGTTCCCAAGCAGCAGAACCGCGCCCAAGCCACCC 300  
Db 241 TTCCCCAGGGAAGGCGCGAGTTCCCAAGCAGCAGAACCGCGCCCAAGCCACCC 300  
Qy 301 AGCGCGAGCTGCAAGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGAG 360  
Db 301 AGCGCGAGCTGCAAGGTGCGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGAG 360  
Qy 361 GGCACTCTGAATCTTCCCGCAGATCACCTGTGGAGAGCGCCCTCTGGTAGCATCAAGGTG 420  
Db 361 GGCACTCTGAATCTTCCCGCAGATCACCTGTGGAGAGCGCCCTCTGGTAGCATCAAGGTG 420  
Qy 421 GGCGGCGAGATCAAGGAGGCGCTGTGAGACCGCGCGCAGCAGACACCGTGTGAGGAG 480  
Db 421 GGCGGCGAGATCAAGGAGGCGCTGTGTGACACCGCGCGCGAGACACCGTGTGAGGAG 480  
Qy 481 ATGAGCCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG 540  
Db 481 ATGAGCCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAG 540  
Qy 541 GTGCGCGAGTACGACAGATCCTGATCGAGATCTTGGCGCAAGAGGCGCATCGGCAACGTTG 600  
Db 541 GTGCGCGAGTACGACAGATCCTGATCGAGATCTTGGCGCAAGAGGCGCATCGGCAACGTTG 600  
Qy 601 CTGATCGGCGCCACCCCGTGAAATCATATCGGCGCAACATGCTGATCGGCGCTTTCATCAAG 660  
Db 601 CTGATCGGCGCCACCCCGTGAAATCATATCGGCGCAACATGCTGATCGGCGCTTTCATCAAG 660  
Qy 661 ACCCTGAATCTTCCCGCAGCCCATCGAGACCGCTGCGGAGCTGAAGCTGAAGCCCGGATG 720  
Db 661 ACCCTGAATCTTCCCGCAGCCCATCGAGACCGCTGCGGAGCTGAAGCTGAAGCCCGGATG 720  
Qy 721 GACGCGCCAAAGGTGAAGAGTGGCGCTGACACCGCGCAACATGCTGATCGGCGCTTTCATCAAG 780  
Db 721 GACGCGCCAAAGGTGAAGAGTGGCGCTGACACCGCGCAACATGCTGATCGGCGCTTTCATCAAG 780  
Qy 781 ATCTGCGAGGAGATGGAAGAGGCGGCAAGATCACCAAGATCGGCGCGCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGGAAGAGGCGGCAAGATCACCAAGATCGGCGCGCGAGAACCCCTAC 840  
Qy 841 AACACCCCGGTGTTCCGCGCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGGTGTTCCGCGCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Qy 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGATCTTCTGGGAGGTGCGAGCTGGGCGATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGATCTTCTGGGAGGTGCGAGCTGGGCGATCCCCAC 960  
Qy 961 CCGCGCGGCTGAAGAGAGAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACTTC 1020  
Db 961 CCGCGCGGCTGAAGAGAGAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACTTC 1020  
Qy 1021 AGCGTGGCCCTGGAGCAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGGCCCTGGAGCAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGCGCATCCGCTACAGTACAACTGTGTGCCCGCAGGGCTGGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGCGCATCCGCTACAGTACAACTGTGTGCCCGCAGGGCTGGAAGGGCAGC 1140  
Qy 1141 CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGGAGCGCTTTCGCGCGCGCAACCCC 1200  
Db 1141 CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGGAGCGCTTTCGCGCGCGCAACCCC 1200  
Qy 1201 GAGATCGTGTACTACAGGCGCCCTGTATCGTGGCGAGCGACCTGGAGATCGGCGAGCAC 1260  
Db 1201 GAGATCGTGTACTACAGGCGCCCTGTATCGTGGCGAGCGACCTGGAGATCGGCGAGCAC 1260  
Qy 1261 CGCGCCAAAGATCGAGGAGCTGCGCAAGCACTGTGTGGCTGGGCTTTCACCAACCCCGGAC 1320  
Db 1261 CGCGCCAAAGATCGAGGAGCTGCGCAAGCACTGTGTGGCTGGGCTTTCACCAACCCCGGAC 1320

Qy 1321 AAGAAGCACAGAAGGAGCCCCCTTCTCTGTGATGGGCTACGAGCTGCACCCCGCAAG 1380  
Db 1321 AAGAAGCACAGAAGGAGCCCCCTTCTCTGTGATGGGCTACGAGCTGCACCCCGCAAG 1380  
Qy 1381 TGGACCGTGCAGGCCATTCGAGCTGCCCGAAGAGAGCTGGACCGTGAACGACATCCAG 1440  
Db 1381 TGGACCGTGCAGGCCATTCGAGCTGCCCGAAGAGAGCTGGACCGTGAACGACATCCAG 1440  
Qy 1441 AAGCTGTGGGCGAGCTGAATCTGGCCAGCAGATCTACCCCGGCGATCAAGGTTGGCGAG 1500  
Db 1441 AAGCTGTGGGCGAGCTGAATCTGGCCAGCAGATCTACCCCGGCGATCAAGGTTGGCGAG 1500  
Qy 1501 CTGTGCAAGCTGTCTGCGCGCGCCCAAGGCCCTTGACCGACATCTGTGCCCTGACCGAGGAG 1560  
Db 1501 CTGTGCAAGCTGTCTGCGCGCGCCCAAGGCCCTTGACCGACATCTGTGCCCTGACCGAGGAG 1560  
Qy 1561 GCGAGCTGAGCTGGCCGAGAACCGCGAGATCTGCGCGAGCCGTGCAAGGCTGTAC 1620  
Db 1561 GCGAGCTGAGCTGGCCGAGAACCGCGAGATCTGCGCGAGCCGTGCAAGGCTGTAC 1620  
Qy 1621 TAGACCCCGCAGAGGAGCTGTGGCCGAGATCCAGAAGCAGGCGCACGACCTAGTGAC 1680  
Db 1621 TAGACCCCGCAGAGGAGCTGTGGCCGAGATCCAGAAGCAGGCGCACGACCTAGTGAC 1680  
Qy 1681 TACCAGATCTACAGGAGGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATGCGC 1740  
Db 1681 TACCAGATCTACAGGAGGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATGCGC 1740  
Qy 1741 ACCGCCCCAACCAACGAGCTGAAGAGCTGACGAGGCCGTGCAAGAGATGCGCCATGGAG 1800  
Db 1741 ACCGCCCCAACCAACGAGCTGAAGAGCTGACGAGGCCGTGCAAGAGATGCGCCATGGAG 1800  
Qy 1801 AGCATCTGTGATCTGGGCGAGAACCCCAAGTTCGCGCTGCCCATCCAGAGAGGAGACCTGG 1860  
Db 1801 AGCATCTGTGATCTGGGCGAGAACCCCAAGTTCGCGCTGCCCATCCAGAGAGGAGACCTGG 1860  
Qy 1861 GAGACCTGTGTGAGCCGACTACTTGGCAGGCCACCTTGGATCCCGAGTGGGAGTTCTGTAAAC 1920  
Db 1861 GAGACCTGTGTGAGCCGACTACTTGGCAGGCCACCTTGGATCCCGAGTGGGAGTTCTGTAAAC 1920  
Qy 1921 ACCCCCCCTGTGTGAAGCTGTGGTACCAAGTGGAGAGGCCCATCTATCGGCGCGAG 1980  
Db 1921 ACCCCCCCTGTGTGAAGCTGTGGTACCAAGTGGAGAGGCCCATCTATCGGCGCGAG 1980  
Qy 1981 ACCTTCTACGTGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGTACTGTG 2040  
Db 1981 ACCTTCTACGTGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGTACTGTG 2040  
Qy 2041 ACCGACCGGCGCGCGCAAGAGATCGTGAGCTTGACCGGAGACCAACCCAGAGACCGAG 2100  
Db 2041 ACCGACCGGCGCGCGCAAGAGATCGTGAGCTTGACCGGAGACCAACCCAGAGACCGAG 2100  
Qy 2101 CTGACGCCATTCAGCTGGCCCTGCGAGGACAGCGGCGAGGAGTGAACATCTGTGACCGAC 2160  
Db 2101 CTGACGCCATTCAGCTGGCCCTGCGAGGACAGCGGCGAGGAGTGAACATCTGTGACCGAC 2160  
Qy 2161 AGCCAGTACGCGCTGGGCGATCATCCAGGCCCGCGAGAGGAGGAGGAGCTGTG 2220  
Db 2161 AGCCAGTACGCGCTGGGCGATCATCCAGGCCCGCGAGAGGAGGAGGAGGAGCTGTG 2220  
Qy 2221 AACAGATCTACGAGCAGCTGATCAAGAAAGAGAGGTGTACTGTAGCTGGGTGCCCGCC 2280  
Db 2221 AACAGATCTACGAGCAGCTGATCAAGAAAGAGAGGTGTACTGTAGCTGGGTGCCCGCC 2280  
Qy 2281 CACAAGGCGATTCGCGCGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATCCGCAAG 2340  
Db 2281 CACAAGGCGATTCGCGCGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATCCGCAAG 2340  
Qy 2341 GTGCTGTCTTGGAGCGGCGATCGATGGGCGATCGTGATCTACCAAGTACATGGAGCAGCTG 2400  
Db 2341 GTGCTGTCTTGGAGCGGCGATCGATGGGCGATCGTGATCTACCAAGTACATGGAGCAGCTG 2400

QY 2401 TACGTGGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
DB 2401 TACGTGGCAGCGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
QY 2461 TTC 2463  
DB 2461 TTC 2463  
RESULT 2  
AX455914  
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 30 from Patent WO0204493.  
ACCESSION AX455914  
VERSION AX455914.1 GI:21714899  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
1  
REFERENCE  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES  
source  
location/Qualifiers  
1..2469  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975(+)"  
ORIGIN  
Query Match 99.2%; Score 2442.2; DB 6; Length 2469;  
Best Local Similarity 99.6%; Pred. No. 5.4e-246;  
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
QY 1 GTCGAGCCACCATGGCGAGGCCATGAGCCAGGCCACAGGCCAAATCTGTATGAG 60  
DB 1 GTCGAGCCACCATGGCGAGGCCATGAGCCAGGCCACAGGCCAAATCTGTATGAG 60  
QY 61 GCGAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAATGCGGCGAAGAGGC 120  
DB 61 GCGAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAATGCGGCGAAGAGGC 120  
QY 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGGAG 180  
DB 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGGAG 180  
QY 181 GGCACACAGATGAAGGACTGCAACCGAGCCAGGCCAACTTCTCCGCGAGGACCTGGCC 240  
DB 181 GGCACACAGATGAAGGACTGCAACCGAGCCAGGCCAACTTCTCCGCGAGGACCTGGCC 240  
QY 241 TTCCCGCAGGCAAGCCCGCGAGTTCCTCCAGCGAGCAAGCCGCGCAAGCCCGACC 300  
DB 241 TTCCCGCAGGCAAGCCCGCGAGTTCCTCCAGCGAGCAAGCCGCGCAAGCCCGACC 300  
QY 301 AGCCCGAGCTCAGGTGCGCGCGCAACACCCCGCAGCGAGCCGCGCGCGAGCGCCAG 360  
DB 301 AGCCCGAGCTCAGGTGCGCGCGCAACACCCCGCAGCGAGCCGCGCGCGAGCGCCAG 360  
QY 361 GGCACCTTGAATTTCCCGCAGATCACTTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTG 420  
DB 361 GGCACCTTGAATTTCCCGCAGATCACTTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTG 420  
QY 421 GCGCGCCAGATCAAGAGGCCCTGTGTGACACCGCGCGCGAGCACACCTGTGTGAGGAG 480  
DB 421 GCGCGCCAGATCAAGAGGCCCTGTGTGACACCGCGCGCGAGCACACCTGTGTGAGGAG 480  
QY 481 ATGAGCTTCCCGCGCAAGTGAAGCCCAAGATGATGCGCGGCATCGCGGGCTTCATCAAG 540  
DB 481 ATGAGCTTCCCGCGCAAGTGAAGCCCAAGATGATGCGCGGCATCGCGGGCTTCATCAAG 540

QY 541 GTGCGCCAGTACGACAGATCTGTATCGAGATCTCGGCGAAGAGGCCATCGGACCGGTG 600  
DB 541 GTGCGCCAGTACGACAGATCTGTATCGAGATCTCGGCGAAGAGGCCATCGGACCGGTG 600  
QY 601 CTGATCGCGCCCGCCACCGCGGTGAATCATCTCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
DB 601 CTGATCGCGCCCGCCACCGCGGTGAATCATCTCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAATTTCCCGCATCAGCCCATCGAGACCGTGTGAACTGTAAGCCCGGCGCATG 720  
DB 661 ACCCTGAATTTCCCGCATCAGCCCATCGAGACCGTGTGAACTGTAAGCCCGGCGCATG 720  
QY 721 GACGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGGC 780  
DB 721 GACGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGGC 780  
QY 781 ATCTCGGAGGAGATGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGAGAACCCCTAC 840  
DB 781 ATCTCGGAGGAGATGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGAGAACCCCTAC 840  
QY 841 AACACCCCGGTTCGCGCATCAAGAGAGGAGACACCAAGTGGCGCAAGCTGTGGAC 900  
DB 841 AACACCCCGGTTCGCGCATCAAGAGAGGAGACACCAAGTGGCGCAAGCTGTGGAC 900  
QY 901 TTCCGCGAGCTGAACAAAGGCCACCCAGGACTTCTGGGAGGTGCAGTGGGCATCCCGCC 960  
DB 901 TTCCGCGAGCTGAACAAAGGCCACCCAGGACTTCTGGGAGGTGCAGTGGGCATCCCGCC 960  
QY 961 CCGCGCGGCTGAAAGAGAGAGCGTGCAGCGGTGGAGCTGGGCGAGCGCTACTTTC 1020  
DB 961 CCGCGCGGCTGAAAGAGAGAGCGTGCAGCGGTGGAGCTGGGCGAGCGCTACTTTC 1020  
QY 1021 AGCTGCGCCCTGGAGAGGACTTTCGCGAAGTACACCGCTTTCACCATCCCGAGCTCAAC 1080  
DB 1021 AGCTGCGCCCTGGAGAGGACTTTCGCGAAGTACACCGCTTTCACCATCCCGAGCTCAAC 1080  
QY 1081 AACGAGACCCCGCGCATCCGCTACAGTACAAAGTGTCTGCCCGAGGCTGGAAGGCGAGC 1140  
DB 1081 AACGAGACCCCGCGCATCCGCTACAGTACAAAGTGTCTGCCCGAGGCTGGAAGGCGAGC 1140  
QY 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCGCAACCCC 1200  
DB 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCGCAACCCC 1200  
QY 1201 GAGATCGTGTACTACCA-----GGCCCCCTGTAGCTGGGCGAGCGACTGTGAGATCGGC 1254  
DB 1201 GAGATCGTGTACTACCA-----GGCCCCCTGTAGCTGGGCGAGCGACTGTGAGATCGGC 1254  
QY 1255 CAGCACCGCGCCAGATCGAGAGCTGCGAAGCACTGTGCGCTGGGGCTTCAACGAC 1314  
DB 1255 CAGCACCGCGCCAGATCGAGAGCTGCGAAGCACTGTGCGCTGGGGCTTCAACGAC 1314  
QY 1315 CCCGAGAGAGACCCAGAGGAGCCCTTCTGTGATGGGCTAGAGCTGACGCTGACCC 1374  
DB 1315 CCCGAGAGAGACCCAGAGGAGCCCTTCTGTGATGGGCTAGAGCTGACGCTGACCC 1374  
QY 1375 GACAAAGTGGAGCCGTGAGCCCATCGAGCTGCCCGAGAGAGAGCTGGAACCGTGAACGAC 1434  
DB 1375 GACAAAGTGGAGCCGTGAGCCCATCGAGCTGCCCGAGAGAGAGCTGGAACCGTGAACGAC 1434  
QY 1435 ATCCGAAAGCTGTGGGCAAGCTGAACTGGGCGAGCAGATCTACCCCGGCTCAAGGTG 1494  
DB 1435 ATCCGAAAGCTGTGGGCAAGCTGAACTGGGCGAGCAGATCTACCCCGGCTCAAGGTG 1494  
QY 1495 CGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCCCTGACCGATCGTGCCTGACCC 1554  
DB 1495 CGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCCCTGACCGATCGTGCCTGACCC 1554  
QY 1555 GAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCGTGACGCGC 1614  
DB 1555 GAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCGTGACGCGC 1614  
QY 1615 GTGTACTACGACCCCGAGAGGAACTGTGTGGCGGAGATTCAGAGAGAGGCGCAACGAC 1674  
DB 1615 GTGTACTACGACCCCGAGAGGAACTGTGTGGCGGAGATTCAGAGAGAGGCGCAACGAC 1674

|||||  
Db 1621 GTGTACTACGACCCAGCAAGACCTGGTGGCCGAGATCCAGAAGAGGCGCACGACGAG 1680  
Qy 1675 TGGACCTACCGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTAGCGCAAG 1734  
Db 1681 TGGACCTACCGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTAGCGCAAG 1740  
Qy 1735 ATGGCACCAGCCACACCAAGAGTGAAGCAGCTGACCGAGGCGGTGCGAAGATCGCC 1794  
Db 1741 ATGGCACCAGCCACACCAAGAGTGAAGCAGCTGACCGAGGCGGTGCGAAGATCGCC 1800  
Qy 1795 ATGAGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTCCGCTCCGCTCCATCCAGAGGAG 1854  
Db 1801 ATGAGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTCCGCTCCGCTCCATCCAGAGGAG 1860  
Qy 1855 ACCTGGAGAGCTGGTGAACCGAATCTGCGAGCGCCACCTGGATCCCGAGTGGGAGTTC 1914  
Db 1861 ACCTGGAGAGCTGGTGAACCGAATCTGCGAGCGCCACCTGGATCCCGAGTGGGAGTTC 1920  
Qy 1915 GTGAACACCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGC 1974  
Db 1921 GTGAACACCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGC 1980  
Qy 1975 GCCAGACCTTCTACTGTGAGCGCGCCGCAACCCGAGACCAAGATCGGCAAGGCCCGC 2034  
Db 1981 GCCAGACCTTCTACTGTGAGCGCGCCGCAACCCGAGACCAAGATCGGCAAGGCCCGC 2040  
Qy 2035 TACGTGACCGACCCGGGCGCGAGAGATCTGAGCTGACCGAGACCAACAGAG 2094  
Db 2041 TACGTGACCGACCCGGGCGCGAGAGATCTGAGCTGACCGAGACCAACAGAG 2100  
Qy 2095 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGAGCAGACAGCGCAGAGTGAACATCGTG 2154  
Db 2101 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGAGCAGACAGCGCAGAGTGAACATCGTG 2160  
Qy 2155 ACCGAGACGCAAGTACGGCCCTGGGCGATCATCAGCGCCAGCGCCGCAAGAGCGAGCGAG 2214  
Db 2161 ACCGAGACGCAAGTACGGCCCTGGGCGATCATCAGCGCCAGCGCCGCAAGAGCGAGCGAG 2220  
Qy 2215 CTGTGAACAGATCATCGACAGCTGATCAAGAGGAGAGGTGACCTGAGCTGGGTG 2274  
Db 2221 CTGTGAACAGATCATCGACAGCTGATCAAGAGGAGAGGTGACCTGAGCTGGGTG 2280  
Qy 2275 CCCCACACAGGCGATCGGCGGCAAGCAGCAGATCGCAGAGCTGTGAGCAAGGCGATC 2334  
Db 2281 CCCCACACAGGCGATCGGCGGCAAGCAGCAGATCGCAGAGCTGTGAGCAAGGCGATC 2340  
Qy 2335 CGCAAGGTGCTGTTCTTGGACGGCATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2394  
Db 2341 CGCAAGGTGCTGTTCTTGGACGGCATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2400  
Qy 2395 GACCTGACGTGGGAGCGCGGCGCTTAGATGATGAATAAAGCTTCCCGGGCTAGCAC 2454  
Db 2401 GACCTGACGTGGGAGCGCGGCGCTTAGATGATGAATAAAGCTTCCCGGGCTAGCAC 2460  
Qy 2455 GGTGAATTC 2463  
Db 2461 GGTGAATTC 2469

RESULT 3  
LOCUS AX455916  
DEFINITION Sequence 32 from Patent WO0204493.  
ACCESSION AX455916  
VERSION AX455916.1 GI:21714901  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
other sequences; artificial sequences.  
REFERENCE  
AUTHORS zur Megede, J.; Barnett, S.W.; Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,

polypeptides and uses thereof  
Patent: WO 0204493-A 32 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES  
source Location/Qualifiers  
1. .2457  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YMM"  
ORIGIN  
Query Match 98.9%; Score 2436.2; DB 6; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 2.3e-245;  
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
Qy 1 GTGACGCCACCATGGCGGAGGCGCATGAGCGAGCCACAGCGCCCAACATCTCTGATCGAG 60  
Db 1 GTGACGCCACCATGGCGGAGGCGCATGAGCGAGCCACAGCGCCCAACATCTCTGATCGAG 60  
Qy 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGC 120  
Db 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGC 120  
Qy 121 CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAG 180  
Db 121 CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAG 180  
Qy 181 GGCCACAGATGAAGGATGCAACGAGCGCGAGCCCAACTTCTCCGAGGAGCTGGCC 240  
Db 181 GGCCACAGATGAAGGATGCAACGAGCGCGAGCCCAACTTCTCCGAGGAGCTGGCC 240  
Qy 241 TTCCCCCAGGCGAGGCGCGAGTTCCCCAGCGAGCAGAAACCGCGCCCAACAGCCCCACC 300  
Db 241 TTCCCCCAGGCGAGGCGCGAGTTCCCCAGCGAGCAGAAACCGCGCCCAACAGCCCCACC 300  
Qy 301 AGCGCGAGCTGCAAGTGTGCGCGGCAACCCCCGCGAGCGGCGCGCGCGAGCGCGAG 360  
Db 301 AGCGCGAGCTGCAAGTGTGCGCGGCAACCCCCGCGAGCGGCGCGCGCGAGCGCGAG 360  
Qy 361 GGCACCTTGAACCTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG 420  
Db 361 GGCACCTTGAACCTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTG 420  
Qy 421 GGCGGCGAGATCAAGGAGGCGCTGTGGAACCGCGCGCGAGCAGACCGTGTCTGGAGAG 480  
Db 421 GGCGGCGAGATCAAGGAGGCGCTGTGGAACCGCGCGCGAGCAGACCGTGTCTGGAGAG 480  
Qy 481 ATGAGCTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGCGCATTCGGCGGCTTCATCAAG 540  
Db 481 ATGAGCTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGCGCATTCGGCGGCTTCATCAAG 540  
Qy 541 GTGCGCGAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTG 600  
Db 541 GTGCGCGAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTG 600  
Qy 601 CTGATCGGCGCCACCCCGGTGAACATCATCGGCGCGCAACATGTGACCCAGCTGGGTGC 660  
Db 601 CTGATCGGCGCCACCCCGGTGAACATCATCGGCGCGCAACATGTGACCCAGCTGGGTGC 660  
Qy 661 ACCCTGAATCTCCCATCAGCCCCATCGAGACCGCTGCGCGTGAAGCTGAAGCCCGGATG 720  
Db 661 ACCCTGAATCTCCCATCAGCCCCATCGAGACCGCTGCGCGTGAAGCTGAAGCCCGGATG 720  
Qy 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTCAGCGCC 780  
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTCAGCGCC 780  
Qy 781 ATCTCGAGGAGATGGAGAGGAGGCAAGATCAACAGATCGGCCCGCGAGAACCCCTTAC 840  
Db 781 ATCTCGAGGAGATGGAGAGGAGGCAAGATCAACAGATCGGCCCGCGAGAACCCCTTAC 840  
Qy 841 AACACCCCGTGTTCATCAAGAGAGAGCAAGATCAAGTGGCGCGAGCTGGTGGAC 900  
Db 841 AACACCCCGTGTTCATCAAGAGAGAGCAAGATCAAGTGGCGCGAGCTGGTGGAC 900

841	Db	841	AACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCAACAAAGTGGCGCAAGCTGGTGGAC	900
901	Qy	901	TTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCCAC	960
	Db	901	TTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCCAC	960
	Qy	961	CCGCGCGGCTGAAGAAAGAAAGAGAGCGTGACCGTGTGGAAGTGGGCGGACGCTACTTC	1020
961	Db	961	CCGCGCGGCTGAAGAAAGAAAGAGAGCGTGACCGTGTGGAAGTGGGCGGACGCTACTTC	1020
	Qy	1021	AGCGTGCCCTGGACGAGGACTTCGCGAAGTACACCGCTTCACCATCCCGACGATCAAC	1080
	Db	1021	AGCGTGCCCTGGACGAGGACTTCGCGAAGTACACCGCTTCACCATCCCGACGATCAAC	1080
1081	Qy	1081	AAAGAGACCCCGGCATCCGCTACCACTACAAACGTGTGCCCCCAGGGCTGGAAGGGCAGC	1140
	Db	1081	AAAGAGACCCCGGCATCCGCTACCACTACAAACGTGTGCCCCCAGGGCTGGAAGGGCAGC	1140
	Qy	1141	CCGAGCATCTTCAGAGAGGATGACCAAGATCTCGAGCGCTTCGCGCGCGCAACCCC	1200
1141	Db	1141	CCGAGCATCTTCAGAGAGGATGACCAAGATCTCGAGCGCTTCGCGCGCGCAACCCC	1200
	Qy	1201	GAGATCGTGATCTACAGGCCCCCTGTACGTGGGACGACCTGGAGATCGGCCAGCAC	1260
	Db	1201	GAGATCGTGATCTACAGGCCCCCTGTACGTGGGACGACCTGGAGATCGGCCAGCAC	1260
1261	Qy	1261	CGGCGCAAGATCAGAGGAGTGGCAAGCACCTGCTGCGCTGGGCTTCACACCCCGCAC	1320
	Db	1261	CGGCGCAAGATCAGAGGAGTGGCAAGCACCTGCTGCGCTGGGCTTCACACCCCGCAC	1320
	Qy	1321	AAGAAGCACAGAAAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCACCCCGACAAG	1380
1321	Db	1321	AAGAAGCACAGAAAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCACCCCGACAAG	1380
	Qy	1381	TGGACCGTGAGCCCATCGAGCTGCCCGAAGAGGAGAGCTGGACCGTGGAACGATCCAG	1440
	Db	1375	TGGACCGTGAGCCCATCGAGCTGCCCGAAGAGGAGAGCTGGACCGTGGAACGATCCAG	1434
1441	Qy	1441	AAGCTGCTGGCGAAGCTGAACCTGGGCGAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG	1500
	Db	1435	AAGCTGCTGGCGAAGCTGAACCTGGGCGAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG	1494
	Qy	1501	CTGTGAAGCTGTGCGCGCGCAAGGCCCTGACCGACATCTGTGCCCTTGACCCGAGGAG	1560
1495	Db	1495	CTGTGAAGCTGTGCGCGCGCAAGGCCCTGACCGACATCTGTGCCCTTGACCCGAGGAG	1554
	Qy	1561	GCGAGCTGAGCTGGCGGAGAACCGGCGAGATCTTGGCGAGCCGCTGACGGGTGTATC	1620
	Db	1555	GCGAGCTGAGCTGGCGGAGAACCGGCGAGATCTTGGCGAGCCGCTGACGGGTGTATC	1614
1621	Qy	1621	TACGACCCAGCAAGGACCTGTGGCGGAGATCCAGAAGCAGGGCCACGACCACTGAGACC	1680
	Db	1615	TACGACCCAGCAAGGACCTGTGGCGGAGATCCAGAAGCAGGGCCACGACCACTGAGACC	1674
	Qy	1681	TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAAGCCGGCAAGTACGCCAAGATGCGC	1740
1675	Db	1675	TACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAAGCCGGCAAGTACGCCAAGATGCGC	1734
	Qy	1741	ACGCGCCACACCAACGACGTGAAGAGCTGACCGAGGCGGTGACAGAGATGTCCTATGGAG	1800
	Db	1735	ACGCGCCACACCAACGACGTGAAGAGCTGACCGAGGCGGTGACAGAGATGTCCTATGGAG	1794
1801	Qy	1801	AGCATCGTGATCTGGGGCAAGCCCCCAAGTTCGCTGCCCATCCAGAGGAGACTGG	1860
	Db	1795	AGCATCGTGATCTGGGGCAAGCCCCCAAGTTCGCTGCCCATCCAGAGGAGACTGG	1854
	Qy	1861	GAGACCTGTGTGACCCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC	1920
1855	Db	1855	GAGACCTGTGTGACCCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC	1914
	Qy	1921	ACCCCGCCCTGTGTGAAGCTGTGTACCACTGGAGAGGAGGCCATCATCGCGCGCAG	1980
	Db	1915	ACCCCGCCCTGTGTGAAGCTGTGTACCACTGGAGAGGAGGCCATCATCGCGCGCAG	1974

Qy	1981	ACCTTCTACGTGGAGCGCGCGCCGCAACCGCGAGACCAAGATTCGGCAAGGCGCGCTACGTG	2040
Db	1975	ACCTTCTACGTGGAGCGCGCGCCGCAACCGCGAGACCAAGATTCGGCAAGGCGCGCTACGTG	2034
Qy	2041	ACCGACCGGGCGCGGACAGAAAGATCGTGGAGCTGACCGAGACCAACCAAGAAAGACCGAG	2100
Db	2035	ACCGACCGGGCGCGGACAGAAAGATCGTGGAGCTGACCGAGACCAACCAAGAAAGACCGAG	2094
Qy	2101	CTGCAGGCGCATTCAGCTGGCCCTTGCAAGGACAGCGGCGAGCGAGGTGAACATCTGTTGACCGAC	2160
Db	2095	CTGCAGGCGCATTCAGCTGGCCCTTGCAAGGACAGCGGCGAGCGAGGTGAACATCTGTTGACCGAC	2154
Qy	2161	AGCCAGTACGCGCTTGGGCGCATCATCCAGGCCGACGCCGACAAAGACGAGACCGAGCTGGTG	2220
Db	2155	AGCCAGTACGCGCTTGGGCGCATCATCCAGGCCGACGCCGACAAAGACGAGACCGAGCTGGTG	2214
Qy	2221	AACAGATCATTCAGAGAGCTGATCAAGAGGAGAGGTGTACTGTAGCTGGGTGCCCGCC	2280
Db	2215	AACAGATCATTCAGAGAGCTGATCAAGAGGAGAGGTGTACTGTAGCTGGGTGCCCGCC	2274
Qy	2281	CACAAGGGCATCGCGGCGCAACGAGCAGATCGACAAGCTGGTGAAGAGGCGCATCCCGAAG	2340
Db	2275	CACAAGGGCATCGCGGCGCAACGAGCAGATCGACAAGCTGGTGAAGAGGCGCATCCCGAAG	2334
Qy	2341	GTGCTGTTCTTGAACGCGCATTCGATGGCGGCATCGTGATCTTACAGTACATGGAACGACCTG	2400
Db	2335	GTGCTGTTCTTGAACGCGCATTCGATGGCGGCATCGTGATCTTACAGTACATGGAACGACCTG	2394
Qy	2401	TACGTGGGACGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCGTAGCACCGGTGAA	2460
Db	2395	TACGTGGGACGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCGTAGCACCGGTGAA	2454
Qy	2461	TTC 2463	
Db	2455	TTC 2457	

RESULT 4	BD263704	2306 bp	DNA	linear	PAT 17-JUL-2003			
LOCUS	BD263704	Improved expression of HIV polypeptides and production of virus-like particles.						
DEFINITION	BD263704	virus-like particles.						
ACCESSION	BD263704.1	GI:33073472						
VERSION	BD263704.1	GI:33073472						
KEYWORDS	JP 200253124-A/71.							
SOURCE	synthetic construct							
ORGANISM	synthetic construct							
REFERENCE	other sequences; artificial sequences.							
AUTHORS	1 (bases 1 to 2306)							
	Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.							
TITLE	Improved expression of HIV polypeptides and production of virus-like particles							
JOURNAL	Patent: JP 200253124-A 71 08-OCT-2002;							
	CHIRON CORP							
COMMENT	OS Artificial Sequence							
	PN JP 200253124-A/71							
	PD 08-OCT-2002							
	PF 30-DEC-1999 JP 2000591193							
	PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI							
	SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI							
	KARIN HARTOG,							
	PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC							
	C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC							
	C12N5/10,							
	PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC							
	Description of Artificial Sequence: FS(-) protmod.RTopt.YM FH							
	Key							
	FT source	Location/Qualifiers						
	1..2306							
FEATURES	FT	/organism='Artificial Sequence'.						
	Location/Qualifiers							
	1..2306							
	source							

/organism="synthetic construct"			
/mol_type="genomic DNA"			
/db_xref="taxon:32630"			
ORIGIN			
Query Match		83.1%; Score 2046; DB 6; Length 2306;	
Best Local Similarity		93.6%; Pred. No. 1.le-204;	
Matches 2159; Conservative		0; Mismatches 135; Indels 12; Gaps 2;	
Qy	170	GCGGCAAGGAGGCGCACGAGATGAAGACTGCAACGAGCGCCAGCGCAACTTCTTCGCGG	229
Dd	1	GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAAATTTCTTCGCGG	60
Qy	230	AGGACCTGGGCTTCCCGCAGGCGAAGCCCGCGAGTTCCCGAGGCGGCGGCGGCGGCGCA	289
Dd	61	AGGACCTGGGCTTCTTCGAGGCGAAGCCCGCGAGTTTCAGCAGCGAGCAGCACCGCGCCA	120
Qy	290	ACAGCCCGCACAGCCGCGAGCTGAGGTGCGCGCG-----ACAACCCCGCAGCGAGG	343
Dd	121	ACAGCCCGCACCGCGCGAGCTGAGGTGCGCGCGCGCGGCGGCGGCGGCGGCGGCGG	180
Qy	344	CCGCGCGCGAGCGCGAGGCGACCTCG-----AACTTCCCGCAGATCACCTGTGGCAGC	397
Dd	181	CCGCGCGCGAGCGCGGCGGCGACCTGTGAGCTTCAACTTCCCGCAGATCACCTGTGGCAGC	240
Qy	398	GCCCCCTGGTGAGCATCAAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	457
Dd	241	GCCCCCTGGTGAGCATCAAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	300
Qy	458	CCGACGACACCGTCTGGAGGAGTGAAGCTTCCCGGCAAGTGAAGCGGCGGCGGCGGCGG	517
Dd	301	CCGACGACACCGTCTGGAGGAGTGAAGCTTCCCGGCAAGTGAAGCGGCGGCGGCGGCGG	360
Qy	518	GCGGCATCGGCGGCTTCAATCAAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	577
Dd	361	GCGGCAATCGGCGGCTTCAATCAAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	420
Qy	578	GCAAGAGGCGCATCGGCGACCGTGTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCA	637
Dd	421	GCAAGAGGCGCATCGGCGACCGTGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCA	480
Qy	638	ACATGTGACCCAGCTGGGCTGCACCTGAATTTCCCGCATCAGCGGCGGCGGCGGCGGCGG	697
Dd	481	ACCTGTGACCCAGATCGGCTGCACCTGAATTTCCCGCATCAGCGGCGGCGGCGGCGGCGG	540
Qy	698	CCGTGAAGCTGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	757
Dd	541	CCGTGAAGCTGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	600
Qy	758	AGAAGATCAAGGCGGCTGACCGCATCTCGAGGAGATGGAGAGGCGGCGGCGGCGGCGGCGG	817
Dd	601	AGAAGATCAAGGCGGCTGACCGCATCTCGAGGAGATGGAGAGGCGGCGGCGGCGGCGGCGG	660
Qy	818	AGATCGGCGCGGAGAACCCCTTCAACACCCCGCTGTTCGCGCATCAAGAGAGGAGGAGGAGCA	877
Dd	661	AGATCGGCGCGGAGAACCCCTTCAACACCCCGCTGTTCGCGCATCAAGAGAGGAGGAGGAGCA	720
Qy	878	CCAAAGTGGCGAAGCTGGTGAAGTTTCGCGAGCTTGAACAAAGCGCACCCAGGACTTCTGGG	937
Dd	721	CCAAAGTGGCGAAGCTGGTGAAGTTTCGCGAGCTTGAACAAAGCGCACCCAGGACTTCTGGG	780
Qy	938	AGGTGACAGTGGGCGATCCCGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	997
Dd	781	AGGTGACAGTGGGCGATCCCGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	840
Qy	998	TGACGCTGGGCGAGCGCTTACTTACGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1057
Dd	841	TGACGCTGGGCGAGCGCTTACTTACGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	900
Qy	1058	CCTTACCATCCCGAGCATCAACAAAGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1117
Dd	901	CCTTACCATCCCGAGCATCAACAAAGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	960
Qy	1118	TGCCCCAGGCGTGAAGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1177
Dd	961	TGCCCCAGGCGTGAAGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1020
Qy	1178	AGCCCTTCCGCGCGCGCAACCCCGAGATCGTGTATCTACAGGCGGCGGCGGCGGCGGCGGCGG	1237
Dd	1021	AGCCCTTCCGCGCGCGCAACCCCGAGATCGTGTATCTACAGGCGGCGGCGGCGGCGGCGGCGG	1080
Qy	1238	GCGACCTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1297
Dd	1081	GCGACCTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1140
Qy	1298	GCTGGGCGTTTCAACACCCCGCAAGAGCAACAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGG	1357
Dd	1141	GCTGGGCGTTTCAACACCCCGCAAGAGCAACAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGG	1200
Qy	1358	GCTACGAGCTGCACCCCGCAAGAGTGAACCGCTGACGCCCATCTATGCTGCGCGAGAGGAGCA	1417
Dd	1201	GCTACGAGCTGCACCCCGCAAGAGTGAACCGCTGACGCCCATCTATGCTGCGCGAGAGGAGCA	1260
Qy	1418	GCTGGACCGTGAAGCGATCCAGAGCTGGTGGGCGAAGCTGAACTGGGGCGCAGCCAGATCT	1477
Dd	1261	GCTGGACCGTGAAGCGATCCAGAGCTGGTGGGCGAAGCTGAACTGGGGCGCAGCCAGATCT	1320
Qy	1478	ACCCCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1537
Dd	1321	ACCCCGCATCAAGGTGAAGCGAGCTGTGCAAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1380
Qy	1538	ACATCTGTCGCCCTGACCGAGGAGGCGGAGCTGGAGCTGGCGCGGAGAAACCGGAGATCTGCG	1597
Dd	1381	AGGTGATTCGCCCTGACCGAGGAGGCGGAGCTGGAGCTGGCGCGGAGAAACCGGAGATCTGCGA	1440
Qy	1598	GCGAGCCGCTGCAAGCGCTGTACTACGACCCCGCAAGAGGAGCTGGTGGCGGCGGAGATCCAGA	1657
Dd	1441	AGGAGCCGCTGCAAGCGCTGTACTACGACCCCGCAAGAGGAGCTGGTGGCGGCGGAGATCCAGA	1500
Qy	1658	AGCAGGCGCCACGACAGTGGAGCTTACAGATCTTACAGAGGCGGCTTCAAGAACCTGAAAGA	1717
Dd	1501	AGCAGGCGCCACGAGTGGAGCTTACAGATCTTACAGAGGCGGCTTCAAGAACCTGAAAGA	1560
Qy	1718	CCGCGAAGTACGCGCAAGATGGCGACCGCGCCACACCAACGAGCTGAAGAGCTGACCGAGG	1777
Dd	1561	CCGCGAAGTACGCGCGCATGCGCGGCGGCGCCACCAACGAGCTGAAGAGCTGACCGAGG	1620
Qy	1778	CCGTGCAAGAGATCGCCATGGAGAGCATCGTGTATCTGGGCGGCAAGACCCCAAGTTCCGCC	1837
Dd	1621	CCGTGCAAGAGTGAAGCAGCAGAGAGCATCGTGTATCTGGGCGGCAAGATCCCCNAGTTCAAGC	1680
Qy	1838	TGCCCATCCAGAGGAGACCTGGGAGACCTGGTGGAGCCGAGCTACTGGCAGGCGCACCTGGA	1897
Dd	1681	TGCCCATCCAGAGGAGACCTGGGAGGCTGGTGGATGGAGTACTGGCAGGCGCACCTGGA	1740
Qy	1898	TCCCGGAGTGGGAGTTCTGTGAACACCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1957
Dd	1741	TCCCGGAGTGGGAGTTCTGTGAACACCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1800
Qy	1958	AGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG	2017
Dd	1801	AGGAGCCCATCATCGTGGCGCGGAGACCTTCTACGTGAGCGGCGGCGGCGGCGGCGGCGGCGG	1860
Qy	2018	AGATCGGCAAGGCGCGCTACGTGACCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	2077
Dd	1861	AGCTGGGCAAGGCGCGCTACGTGACCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1920
Qy	2078	AGACCAACCAAGAGACCGAGCTGACGCGCATTCAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGG	2137
Dd	1921	ACACCAACCAAGAGACCGAGCTGACGCGCATTCACCTGGGCGGCGGCGGCGGCGGCGGCGGCGG	1980
Qy	2138	GCGAGGTGAACATCGTGCACCGAGCGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	2197
Dd	1981	TGGAGGTGAACATCGTGCACCGAGCGAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	2040
Qy	2198	ACAAGAGCGAGAGCGGAGCTGGTGAACCGAGATCATTCGAGGAGCTGATCAAGAGGAGGAGG	2257

```
Db 2041 ACAAGAGCCAGAGCGAGCTGGTGAGCCCATCATCGAGCGCTGATCAAGAAGGAGAGG 2100
Qy 2258 TGTACTGAGCTGGGTGCGCCGCCACAAAGGGCATCGCGCGCAACGAGCAGATCGACAAGC 2317
Db 2101 TGTACTGCGCTGGGTGCGCCGCCACAAAGGGCATCGCGCGCAACGAGCAGGTGGACAAGC 2160
Qy 2318 TGGTAGCAGAGGCATCCGCAAGGTGCTGTTCTTCTGGAAGGATCGATGCGCGCATCGTGA 2377
Db 2161 TGGTAGCGCCGCGCATCCGCAAGGTGCTGTTCTTCTGGAAGGATCGATGCGCGCATCGTGA 2220
Qy 2378 TCTACAGTACATGACGACCTGTAGTGGGAGCGGGGCGGCTAGGATCGATTAAAGC 2437
Db 2221 TCTACAGTACATGACGACCTGTAGTGGGAGCGGGGCGGCTAGGATCGATTAAAGC 2280
Qy 2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db 2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306

RESULT 5
LOCUS Q0870574 2306 bp DNA linear PAT 13-SEP-2004
DEFINITION Sequence 82 from Patent EP1433851.
ACCESSION Q0870574
VERSION Q0870574.1 GI:52000090
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
AUTHORS Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL Patent: EP 1433851-A 82 30-JUN-2004;
CHIRON CORPORATION (US)
FEATURES
source
location/Qualifiers
1..2306
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description of Artificial Sequence:
FS(-).protmod.RTopt.YM"
ORIGIN
Query Match 83.1%; Score 2046; DB 6; Length 2306;
Best Local Similarity 93.6%; Pred. No. 1.1e-204;
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

Qy 170 GCGCAAGAGGGCCACAGATGAAGGACTGCACCGAGCGCCAGGCGCAACTTTCTTCGCG 229
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCCTGAGAGACAGGCTAATTCTTTCGCG 60
Qy 230 AGGACTGGCTTCCCGCAGGCAAGCGCGGAGTTCCCGAGAGCAGCAACCGCGCA 289
Db 61 AGGACTGGCTTCCCGCAGGCAAGCGCGGAGTTCCCGAGAGCAGCAACCGCGCA 120
Qy 290 ACAGCCCAACAGCGCGAGCTGCAAGGTGCGCGGCG-----ACAACCCCGCAGCGAGG 343
Db 121 ACAGCCCAACAGCGCGAGCTGCAAGGTGCGCGGCGGCGAGAACACAGCTGAGCGAGG 180
Qy 344 CCGGCGCGAGCGCCAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGC 397
Db 181 CCGGCGCGAGCGCCAGGCGACCGTGAAGTTCAACTTCCCGCAGATCACCTGTGGCAGC 240
Qy 398 GCCCTCTGGTAGCATCAAGGTGGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGC 457
Db 241 GCCCTCTGGTAGCATCAAGGTGGCGCGCAGCTCAAGAGGCGCTGTGGACACCGCGC 300
Qy 458 CCGAGCACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAGGCCAGATGATCG 517
Db 301 CCGAGCACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAGGCCAGATGATCG 360
```

```
Qy 518 GCGGATCGGCGGCTTTCATCAAGGTGCGCAGTACGACAGATCCTGTGATCGAGATCTGCG 577
Db 361 GCGGATCGGCGGCTTTCATCAAGGTGCGCAGTACGACAGATCCTGTGATCGAGATCTGCG 420
Qy 578 GCAAGAAGCCCATCGGCACCGTGTGATCGGCGCCCAACCCCGTGAACATCATCGGCCGCA 637
Db 421 GCCACAGGCCCATCGGCACCGTGTGATCGGCGCCCAACCCCGTGAACATCATCGGCCGCA 480
Qy 638 ACATCTGACCCAGCTGGGCTGCACCCCTGAACTTCCCATCAGCCCGCATCGAGACGGTGC 697
Db 481 ACCTGCTGACCCAGATCGGCTGCACCCCTGAACTTCCCATCAGCCCGCATCGAGACGGTGC 540
Qy 698 CCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGGTGAAGCAGTGGGCCCTTGACCGAGG 757
Db 541 CCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAGGTGAAGCAGTGGGCCCTTGACCGAGG 600
Qy 758 AGAGATCAAGGCCCTGACCGCATCTCGAGGAGATGAGAGAGGAGGCAAGATCACCA 817
Db 601 AGAGATCAAGGCCCTGAGGATCTGACCGAGATGAGAGAGGAGGCAAGATCACCA 660
Qy 818 AGATCGGCCCGAGAACCCCTACACACCCCGCTGTTCCCATCAAGAGAGAGGACAGCA 877
Db 661 AGATCGGCCCGAGAACCCCTACACACCCCGCTGTTCCCATCAAGAGAGAGGACAGCA 720
Qy 878 CCAAGTGGCGCAAGCTGTGTGACTTCCCGGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 937
Db 721 CCAAGTGGCGCAAGCTGTGTGACTTCCCGGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 780
Qy 938 AGGTGACGTGGGCATCCCGCCCGCGCTGAAAGAGAGAGGAGCGTGAACCGTGC 997
Db 781 AGGTGACGTGGGCATCCCGCCCGCGCTGAAAGAGAGAGGAGCGTGAACCGTGC 840
Qy 998 TGGAGCTGGGCGAGCGCTTACTTTCAGCGTGCCTCGAGAGGACTTCCCGCAAGTACACCG 1057
Db 841 TGGAGCTGGGCGAGCGCTTACTTTCAGCGTGCCTCGAGAGGACTTCCCGCAAGTACACCG 900
Qy 1058 CTTTACCATCCCGAGCATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
Db 901 CTTTACCATCCCGAGCATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Qy 1118 TGCCCCAGGCTGGAAGGGAGCGCCAGCATCTTTCAGAGCAGCATGACCCAGAGATCTGCG 1177
Db 961 TGCCCCAGGCTGGAAGGGAGCGCCAGCATCTTTCAGAGCAGCATGACCCAGAGATCTGCG 1020
Qy 1178 AGCCCTTCGCGCGCCGCAACCCCGAGATCTGATCTACAGCGCCCGCTGTACGTGGGCA 1237
Db 1021 AGCCCTTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy 1238 GCGACCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCTGC 1297
Db 1081 GCGACCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGAGCTGCGCGCAGCACTGTGCTGC 1140
Qy 1298 GCTGGGCTTCAACACCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
Db 1141 GCTGGGCTTCAACACCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Qy 1358 GCTACGAGCTGCACCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
Db 1201 GCTACGAGCTGCACCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1418 GCTGAGCGGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGGCGCAGATCT 1477
Db 1261 GCTGAGCGGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGGCGCAGATCT 1320
Qy 1478 ACCCGGAGATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGGCGCGCAGGCGCTGACCG 1537
Db 1321 ACAGCGGATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGGCGCGCAGGCGCTGACCG 1380
Qy 1538 ACATCTGCGCTGACCGAGGAGCGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
Db 1381 AGGTGATCCCCCTGACCGAGGAGCGAGCTGGAGCTGGCGAGAGAGAGAGAGAGAGAGAG 1440
```



Qy 1598 GCGAGCCGTGCGCGGTGTACTACGACCCACAGGACCTGGTGGCCGAGATCCAGA 1657  
Db |||||||  
Qy 1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCACAGGACCTGGTGGCCGAGATCCAGA 1500  
Db |||||||  
Qy 1658 AGCAGGCCACGACGACGAGTGGACCTACAGATCTACAGGAGCCCTTCAAGAACCTGAAGA 1717  
Db |||||||  
Qy 1501 AGCAGGCCACGAGGCCGAGTGGACCTACAGATCTACAGGAGCCCTTCAAGAACCTGAAGA 1560  
Db |||||||  
Qy 1718 CCGGCAAGTACGCAAGATGGCAGCCGCGCCACACCAACGAGCTGAAGAGTGCACGAGG 1777  
Db |||||||  
Qy 1561 CCGGCAAGTACGCGCCGATGCGCGCGCCACACCAACGAGCTGAAGAGTGCACGAGG 1620  
Db |||||||  
Qy 1778 CCGTGCAGAAAGTACGCAAGAGGAGTGGTGTACTGGGGCAAGACCCCAAGTTCGCGC 1837  
Db |||||||  
Qy 1621 CCGTGCAGAAAGTACGCAAGAGGAGTGGTGTACTGGGGCAAGATCCCCAAGTTCAGC 1680  
Db |||||||  
Qy 1838 TGCCCATCCAGAAAGGAGACCTTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897  
Db |||||||  
Qy 1681 TGCCCATCCAGAAAGGAGACCTTGGGAGGCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740  
Db |||||||  
Qy 1898 TCCCGAGTGGGAGTGGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGA 1957  
Db |||||||  
Qy 1741 TCCCGAGTGGGAGTGGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGA 1800  
Db |||||||  
Qy 1958 AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCA 2017  
Db |||||||  
Qy 1801 AGGAGCCCATCGTGGGCGCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCA 1860  
Db |||||||  
Qy 2018 AGATCGGCAAGCCGGCTTACGTGACCGACCGGGGCGCGGAGAGATCGTGAAGCTGACCG 2077  
Db |||||||  
Qy 1861 AGCTGGGCAAGCCGGCTACGTGACCGACCGGGCGCGGAGAGTGGTGAAGCATCGCG 1920  
Db |||||||  
Qy 2078 AGACACCAACAGAGACCGAGCTGACGAGCCATCAGCTGGCCCTGGAGAGACGAGCGCA 2137  
Db |||||||  
Qy 1921 ACACACCAACAGAGACCGAGCTGACGAGCCATCCACCTGGCCCTGGCAGGACAGCGCC 1980  
Db |||||||  
Qy 2138 GCGAGGTGAACATCGTGACCGACAGCAGTACGAGCCCTGGGATCATCAGGCCCGAGCCG 2197  
Db |||||||  
Qy 1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGAGCCCTGGGATCATCAGGCCCGAGCCG 2040  
Db |||||||  
Qy 2198 ACAAGAGCGAGCGAGCTGGTGAACCCAGATCATCGAGAGCTGATCAAGAAGGAGAAGG 2257  
Db |||||||  
Qy 2041 ACAAGAGCGAGCGAGCTGGTGAAGCCAGATCATCGAGAGCTGATCAAGAAGGAGAAGG 2100  
Db |||||||  
Qy 2258 TGTACTGAGCTGGGTGCGCCGCCACAGGCGATCGGGGCAACGAGGAGATCGACAGC 2317  
Db |||||||  
Qy 2101 TGTACTGAGCTGGGTGCGCCGCCACAGGCGATCGGGGCAACGAGGAGTGGACAGC 2160  
Db |||||||  
Qy 2318 TGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTTGGAGCGCATCGATGGCGGATCGTGA 2377  
Db |||||||  
Qy 2161 TGGTGAGCGCGGCATCCGCAAGGTGCTGTTCTTGAACGGCATCGATGGCGGCATCGTGA 2220  
Db |||||||  
Qy 2378 TCTACAGTACATGGAACGACCTGTACGTGGGAGCGCGGCGCCCTAGGATCGATTAAAGC 2437  
Db |||||||  
Qy 2221 TCTACAGTACATGGAACGACCTGTACGTGGGAGCGCGGCGCCCTAGGATCGATTAAAGC 2280  
Db |||||||  
Qy 2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463  
Db |||||||  
Qy 2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306  
Db |||||||

RESULT 6  
LOCUS AR373387  
DEFINITION Sequence 82 from patent US 6602705.  
ACCESSION AR373387  
VERSION AR373387.1 GI:40075490  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2306)  
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.

TITLE Expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: US 6602705-A 82 05-AUG-2003;  
FEATURES Location/Qualifiers  
source 1. .2306  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 83.1%; Score 2046; DB 6; Length 2306;  
Best Local Similarity 93.6%; Pred. No. 1.1e-204;  
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
Qy 170 GCGCAAGGAGGGCCACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGC 229  
Db |||||||  
Qy 230 AGGACCTTGGCTTCCCGAGGCAAGCCCGAGGTTCCCGAGGAGAGCAACCGCGCA 289  
Db |||||||  
Qy 61 AGGACCTTGGCTTCTTCAGGGCAAGGCCCGAGTTTCAGCGAGCAGACCGCGCA 120  
Db |||||||  
Qy 290 ACAGCCCCACAGCGCGAGCTGCAGGTGCGCGCGC-----ACAAACCCCGAGCGAGG 343  
Db |||||||  
Qy 121 ACAGCCCCACCGCGCGAGCTGCAGGTGTGGGGCGCGAGAACAAAGCGCTGAGCGAGG 180  
Db |||||||  
Qy 344 CCGCGCGCGAGCGCGAGGCAACCTG-----AACTTCCCGAGATCACTGTGGGAGC 397  
Db |||||||  
Qy 181 CCGCGCGCGAGCGCGAGGCAACCTGAGCTTCAACTTCCCGAGATCACTGTGGGAGC 240  
Db |||||||  
Qy 398 GCCCCTTGGTGAAGATCAAGTGGCGGCCAGATCAAGGAGGCCCTGCTGAGACACCGCGC 457  
Db |||||||  
Qy 241 GCCCCTTGGTGAACATCAGGATCGCGGCCAGCTCAAGGAGGCGCTGCTGACACCGCGC 300  
Db |||||||  
Qy 458 CCGAGCACCGCTGCTGGAGGAGTGAAGCTGCCCCGCAAGTGAAGCCCAAGATGATCG 517  
Db |||||||  
Qy 301 CCGAGCACCGCTGCTGGAGGAGTGAAGCTGCCCCGCAAGTGAAGCCCAAGATGATCG 360  
Db |||||||  
Qy 518 GCGGCATCGCGCGCTTCATCAAGGTGCGCCAGTACGACAGATCCTGATCGAGATCTGCG 577  
Db |||||||  
Qy 361 GCGGATCGGGGCTTCATCAAGTGGGAGTACGACAGATCCCGCTGAGATCTGCG 420  
Db |||||||  
Qy 578 GCAAGAGGCCATCGGCACCGTGTGATCGGCCCAACCCCGTGAATCATCATCGGCGCA 637  
Db |||||||  
Qy 421 GCCAAGGCCATCGGCACCGTGTGTTGGGCCCAACCCCGTGAATCATCATCGGCGCA 480  
Db |||||||  
Qy 638 ACATGTCACCCAGCTGGGCTGCACCTGAACTTCCCATCAGCCCATCGAGACCTGTC 697  
Db |||||||  
Qy 481 ACCTGCTGACCCAGATCGGCTGCACCTGAACTTCCCATCAGCCCATCGAGACGCTGC 540  
Db |||||||  
Qy 698 CCGTGAAGCTGAAGCCCGCATGGAAGCGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 757  
Db |||||||  
Qy 541 CCGTGAAGCTGAAGCCCGGATGGAAGCGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 600  
Db |||||||  
Qy 758 AGAAGATCAAGGCCCTGACCCCATCTCGAGGAGATGGAAGAGGAGGCAAGATCAACA 817  
Db |||||||  
Qy 601 AGAAGATCAAGGCCCTGTTGGAGATCTGCAACCGAGATGGAAGAGGAGGCAAGATCAACA 660  
Db |||||||  
Qy 818 AGATCGGCCCGAGAACCCCTTACAACCCCGCTGTTGGCCATCAAGAGAGGAGCAGCA 877  
Db |||||||  
Qy 661 AGATCGGCCCGAGAACCCCTTACAACCCCGCTGTTGGCCATCAAGAGAGGAGCAGCA 720  
Db |||||||  
Qy 878 CCAAGTGGCGCAAGCTGGTGGATTTCCGCGAGCTGAACAGCGCAACCCAGGACTTCTGGG 937  
Db |||||||  
Qy 721 CCAAGTGGCGCAAGCTGGTGGATTTCCGCGAGCTGAACAGCGCAACCCAGGACTTCTGGG 780  
Db |||||||  
Qy 938 AGGTGAGCTGGGCATCCCCACCCCGGCTTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGG 997  
Db |||||||  
Qy 781 AGGTGAGCTGGGCATCCCCACCCCGGCTTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGG 840  
Db |||||||  
Qy 998 TGGACGTGGGCGACCGCTACTTTCAGCTGCCCCCTGGAGGAGGAGGAGGAGGAGGAGGAGG 1057  
Db |||||||  
Qy 841 TGGACGTGGGCGACCGCTACTTTCAGCTGCCCCCTGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
Db |||||||



Db 3260 AGTTTGGGAAGAGACAACAACCTCCCTCTCAGAAAGCAGGAGCCGATAGACAAGAACTG 3319  
Qy 365 ---CCCTGAACCTTCCCGCAGATCAACCTGTGGAGCGCCCTGTGTGAGCATCAAGGTGG 421  
Db 3320 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCAGCCCTCTGCATCAATAAGATAG 3379  
Qy 422 GCGGCCAGATCAAGAGGCGCTGTGGACACCGGCGCGAGCACACGCTGTGGAGGAGA 481  
Db 3380 GGGGCCAGCTGAAGAGGCGCTTCTAGACACCGGCGCGAGCACACGCTGTGGAGGAGA 3439  
Qy 482 TGAGCCTGCCCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAGG 541  
Db 3440 TGAACCTGCCCGCGCTGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAGG 3499  
Qy 542 TGGCCAGTACGACCAAGATCTGTATCGAGATCTCGGCAAGAGCCATCGGCACCGTGC 601  
Db 3500 TGGGCCAGTACGACCAAGATCTGTATCGAGATCTCGGCAAGAGCCATCGGCACCGTGC 3559  
Qy 602 TGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCA 661  
Db 3560 TGGTGGGCCCCACCCCGTGAACATCATCGGCGCAACCTGCTGACCCAGATCGGCTGCA 3619  
Qy 662 CCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATGG 721  
Db 3620 CCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATGG 3679  
Qy 722 ACGGCCCCAAGGTGAAGCAGTGGCCCCGTGACCGAGAGAGATCAAGGCCCTGACCGCCA 781  
Db 3680 ACGGCCCCAAGGTGAAGCAGTGGCCCCGTGACCGAGAGAGATCAAGGCCCTGAGTGA 3739  
Qy 782 TCTCGAGGAGATCGAGAGAGGCGCAGATCAACCAAGTGGCGCAAGCTGGTGGACT 841  
Db 3740 TCTGCAACGAGATGGAGAGAGGCGCAGATCAACCAAGTGGCGCAAGCTGGTGGACT 3799  
Qy 842 ACACCCCGCTTTCGCCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 901  
Db 3800 ACACCCCGCTTTCGCCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACT 3859  
Qy 902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGACGTGGGCAATCCCCAAC 961  
Db 3860 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGACGTGGGCAATCCCCAAC 3919  
Qy 962 CCGCGGCTTGAAGAGAGAGAGCTGACCGTGTGAGCGTGGCGAGCGCTACTTCA 1021  
Db 3920 CCGCGGCTTGAAGAGAGAGAGCTGACCGTGTGAGCGTGGCGAGCGCTACTTCA 3979  
Qy 1022 GCGTGCCCTGGACGAGGACTTTCGCCAAGTACACCGCTTCACCATCCCCAGCATCAACA 1081  
Db 3980 GCGTGCCCTGGACAGGACTTTCGCCAAGTACACCGCTTCACCATCCCCAGCATCAACA 4039  
Qy 1082 ACGAGACCCCGGATCCGCTACAGTACAAAGTGTGCCCCAGGGCTGGAAGGCGAGCC 1141  
Db 4040 ACGAGACCCCGGATCCGCTACAGTACAAAGTGTGCCCCAGGGCTGGAAGGCGAGCC 4099  
Qy 1142 CCAGCATTTCCAGAGCAGATGACCAAGATTCCTGGAGCCCTTCGCGCCCGCAACCCCG 1201  
Db 4100 CCGCATCTTCCAGTGCAGCATGACCAAGATTCCTGGAGCCCTTCGCAAGCAGAACCCCG 4159  
Qy 1202 AGATCGTGATCTACCA-----GGCCCCCTGTGACGTGGGCGAGGACCTGGAGATCGGCC 1255  
Db 4160 ACATCGTGATCTACCAAGTACATGAGCCACCTGTGACGTGGGCGAGGACCTGGAGATCGGCC 4219  
Qy 1256 AGCACCGGCGCAAGATCGAGAGCTGCGCAAGACCTGCTCGGCTGGGGCTTTCACACCC 1315  
Db 4220 AGCACCGCACCAAGATCGAGAGCTGCGCCAGCACCTGCTCGGCTGGGGCTTTCACACCC 4279  
Qy 1316 CCGACAGAGACCAAGAGAGGAGCCCTTCTCTGTGATGGGCTACGAGCTGCACCCCG 1375  
Db 4280 CCGACAGAGACCAAGAGAGGAGCCCTTCTCTGTGATGGGCTACGAGCTGCACCCCG 4339  
Qy 1376 ACAAGTGACCGTGCAGGCCATCGAGCTGCCCCGAGAGAGAGCTGGACCGTGAACGACA 1435

Db 4340 ACAAGTGGACGCTGCAGCCCATCGTGTGCTGCCGAGAGAGACAGCTGACCGTGAACGACA 4399  
Qy 1436 TCCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCGACAGAGATCTACCCGGGATCAAGGTGC 1495  
Db 4400 TCCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCGACAGAGATCTACCCGGGATCAAGGTGC 4459  
Qy 1496 GCCAGCTGTGCAAGCTGCTGCGGGCGCCAAAGGCCCTTGACCGACATCGTGCCTGACCG 1555  
Db 4460 GCCAGCTGTGCAAGCTGCTGCGGGCACCAAGGCCCTTGACCGAGGTGTGCTGACCG 4519  
Qy 1556 AGGAGCGGAGCTGGAGTGGCGGAGAAACCGCGAGATCTTGTGCGGAGCCCGTGCACGGCG 1615  
Db 4520 AGGAGCGGAGCTGGAGTGGCGGAGAAACCGCGAGATCTTGAAGGAGCCCGTGCACGGCG 4579  
Qy 1616 TGTACTACGACCCCGACAGGACCTGTGTGGCGGAGATCCAGAGCAGGCGCACACAGT 1675  
Db 4580 TGTACTACGACCCCGACAGGACCTGTATCCCGGAGATCCAGAGCAGGCGCAGGGCCAGT 4639  
Qy 1676 GGACCTACAGATCTACGAGGAGCCCTTCAAGAACCTTGAAGACCGCGCAAGTACGCCAAGA 1735  
Db 4640 GGACCTACAGATCTACGAGGAGCCCTTCAAGAACCTTGAAGACCGCGCAAGTACGCCGCA 4699  
Qy 1736 TGGCCACCGGCCACACAAAGAGAGCTGAAGAGCTGACCGAGGCGCTGCAGAGATCGCCA 1795  
Db 4700 TGAAGGCGGCCACACCAAGAGAGCTGAAGAGCTGACCGAGGCGCTGCAGAGATCGCCA 4759  
Qy 1796 TGGAGAGCATGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGAGAGA 1855  
Db 4760 CCGAGAGCATGTGATCTGGGGCAAGACCCCAAGTTCAAGTTCAAGTCTCCCATCCAGAGGAGA 4819  
Qy 1856 CTTGGGAGAGACTGTGTGACCGACTACTGTGGCAGGCGCACTTGATCCCGAGTGGGAGTTGC 1915  
Db 4820 CTTGGGAGGCTGTGTGACCGAGTACTGTGGCAGGCGCACTTGATCCCGAGTGGGAGTTGC 4879  
Qy 1916 TGAACACCCCGCTGTGTGAAAGCTGTGTAACAGCTGTGAGAGAGGAGCCATCATCGCGC 1975  
Db 4880 TGAACACCCCGCTGTGTGAAAGCTGTGTAACAGCTGTGAGAGAGGAGCCATCATCGCGC 4939  
Qy 1976 CCGAGACTTCTACGTGGAGCGGCGCGCAACCGCGAGAGCCAGATCGGCAAGGCCCGCT 2035  
Db 4940 CCGAGACTTCTACGTGGAGCGGCGCGCAACCGCGAGAGCCAGCTGGGCAAGGCCCGCT 4999  
Qy 2036 ACGTGACCGGAGCGGCGCGCAGAGATCGTGAAGCTTGAAGCCCTGAGACCAACCAAGAGAGA 2095  
Db 5000 ACGTGACCGGAGCGGCGCGCAGAGAGTGTGTGCGCTTGAACCGACACCAACCAAGAGAGA 5059  
Qy 2096 CCGAGCTGACAGGCGCATCCAGCTGTGCGCTTGCAGGACAGCGGCGAGGAGTGAACATCGTGA 2155  
Db 5060 CCGAGCTGACAGGCGCATCCAGCTGTGCGCTTGCAGGACAGCGGCGCTGAGAGTGAACATCGTGA 5119  
Qy 2156 CCGACAGCGAGTACGCGCTGGGCATCATCCAGGCGCGAGCCCGCAAGAGCGAGAGCGAGC 2215  
Db 5120 CCGACAGCGAGTACGCGCTGGGCATCATCCAGGCGCGAGCCCGCAAGAGCGAGAGCGAGC 5179  
Qy 2216 TGGTGAACACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTGTACTGTAGCTGGGTGC 2275  
Db 5180 TGGTGAACACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTGTACTGTGGCTGGGTGC 5239  
Qy 2276 CCGCCCCAAGAGGCGATCGGCGGCAACGAGCAGATTCGACAGCTGTGTGAGCAGAGGCGATCC 2335  
Db 5240 CCGCCCCAAGAGGCGATCGGCGGCAACGAGCAGGTGTGACGCGCTGTGTGAGCGCGCGATCC 5299  
Qy 2336 GCAAGGTGTGTCTTCGTGAGCGGCGATCGATGGGCGCATCGTGTATCTACAGTA 2387  
Db 5300 GCAAGGTGTGTCTTCGTGAGCGGCGATCGACAAAGGCCCGAGGAGGAGCAGAGAA 5351

RESULT 8  
BD263706  
LOCUS  
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.  
ACCESSION BD263706

2312 bp DNA linear PAT 17-JUL-2003



1772 CCAGAGCGCTGCAGAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGT 1831  
1621 CCAGAGCGCTGCAGAGATGAGACACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680  
1832 TCCGCTCTGCCATCCAGAAGGAGACCTGGGAGACCTTGGTGAGACCTACTTGGCAGGCCA 1891  
1681 TCAAGCTGCCATCCAGAAGGAGACCTGGGAGGCTTGGTGATGGTACTTGGCAGGCCA 1740  
1892 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGGTACTGAGC 1951  
1741 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGGTACTGAGC 1800  
1952 TGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTAGCTGGAGCGCGCGCCCAACCGCG 2011  
1801 TGGAGAAGGAGCCCATCTGTGGCGCCGAGACCTTCTAGCTGGAGCGCGCGCCCAACCGCG 1860  
2012 AGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGAGCGCGCGCCGAGAAATCGTAGCC 2071  
1861 AGACCAAGCTGGGCAAGCGCGGCTACGTGACCGACCGAGCGCGCGCCGAGAAATCGTAGCA 1920  
2072 TGACCGAGACACCAACAGAGACCGAGCTGCGAGGCGCATCCAGCTGGCCCTGCAGGACA 2131  
1921 TCGCCGACACCAACAGAGACCGAGCTGCGAGGCGCATCCACCTGGCCCTGCAGGACA 1980  
2132 GCGGACGAGGTGAACATCGTGACCGACGACCGAGTACGCCCTGGGCGATCATCCAGGCC 2191  
1981 GCGGCTTGGAGTGAAATCGTGACCGACGACCGAGTACGCCCTGGGCGATCATCCAGGCC 2040  
2192 AGCCCGCAAGAGCGAGCGAGCTGGTGAAACAGATCATCGAGCAGCTGATCAAGAAGG 2251  
2041 AGCCCGCAAGAGCGAGCGAGCTGGTGAGCGAGATCATCGAGCAGCTGATCAAGAAGG 2100  
2252 AGAAGTGTACCTGAGCTGGTGCCCGCCCAAGAGGCGATCGCGGCGCAAGAGATCG 2311  
2101 AGAAGTGTACCTGGCTGGTGCCCGCCCAAGAGGCGATCGCGGCGCAAGAGATCG 2160  
2312 ACAAGCTGGTGAGCAAGGCGATCGCAAGGTGCTGTTCTTGGAGCGGATCGATCGCGGCA 2371  
2161 ACAAGCTGGTGAGCGCGGCGATCGCAAGGTGCTGTTCTTGGAGCGGATCGATCGCGGCA 2220  
2372 TCGTGATCTACAGTACATGAGACCTGTACGTGGGCGAGCGCGCCCTAGGATCGATT 2431  
2221 TCGTGATCTACAGTACATGAGACCTGTACGTGGGCGAGCGCGCGCCCTAGGATCGATT 2280  
2432 AAAAGCTTCCGGGCTAGCACCGGTGAATTC 2463  
2281 AAAAGCTTCCGGGCTAGCACCGGTGAATTC 2312

RESULT 9  
LOCUS CQ870576 2312 bp DNA linear PAT 13-SEP-2004  
DEFINITION Sequence 84 from Patent EP143851.  
ACCESSION CQ870576  
VERSION CQ870576.1 GI:52000092  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Barnett,S., Zurmege,J., Srivastava,I., Lian,Y., Hartog,K.,  
Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: EP 143851-A 84 30-JUN-2004;  
CHIRON CORPORATION (US)  
FEATURES  
source  
1..2312  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Description of Artificial Sequence:  
FS (-). protmod.RTopt (+)"

ORIGIN  
Query Match 82.2%; Score 2025.2; DB 6; Length 2312;  
Best Local Similarity 93.3%; Pred. No. 1.7e-202;  
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
170 GCAGCAAGAGGAGGCGCCACAGATGAAGGACTGACGCGCAGCGCCAGGCCAACTTCTTCGCGC 229  
1 GCGGCGCGGAAGGACCAATGAAGATTGCATGAGAGCAGGCTAAATTTCTTCGCGC 60  
230 AGGACTTGGCTTCCCGCCAGGCGAAGCCCGCGAGTTCCCGAGCGAGCAGAACCGCGCCA 289  
61 AGGACTTGGCTTCTCTGAGGCGAAGCCCGCGAGTTTCAAGGAGGCGCTGCTCGACACCGCGCCA 120  
290 ACAGCCCCACAGCGCGAGCTGCAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 343  
121 ACAGCCCCACCGCGCGAGCTGCAGTGTGGGCGCGCGAGAACACAGCTTGAGCGAGG 180  
344 CCGGCGCGGAGCGCGAGGCGACCTG-----AACTTCCCGCCAGATCACCTGTGGCAGC 397  
181 CCGGCGCGGAGCGCGAGGCGACCGTGAGCTTCAACTTCCCGCAGATCACCTGTGGCAGC 240  
398 GCGGCTTGGTGAAGATCAAGTGGGCGCGCAGATCAAGGAGGCGCTGCTGGAACCGCGCG 457  
241 GCGGCTTGGTGAAGATCAAGTGGGCGCGCAGCTCAAGGAGGCGCTGCTGGAACCGCGCG 300  
458 CCGACGACACCGTCTGGAGGAGTGAAGCTGCGCGCGAGTGAAGCGCCCAAGATGATCG 517  
301 CCGACGACACCGTCTGGAGGAGTGAAGCTGCGCGCGAGTGAAGCGCCCAAGATGATCG 360  
518 GCGGCTTGGGCGCTTCAATCAAGTGGCGCAGTACGACAGATCCTGATCGAGATCTGCG 577  
361 CCGGCTTGGGCGCTTCAATCAAGTGGCGCAGTACGACAGATCCTGCGGATCTGCG 420  
578 GCAAGAGGCGCATCGGCAACCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 637  
421 GCGCAAGGCGCATCGGCAACCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
638 ACATGTCACCGGAGCTGGGCGTGCACCTGAATCTCCCGCATCAGCCCATCGAGACCGTGC 697  
481 ACCTGTGACCGAGATCGGCTGCACCTGAATCTCCCGCATCAGCCCATCGAGACCGTGC 540  
698 CCGTGAAGCTGAAGCGCGGATGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757  
541 CCGTGAAGCTGAAGCGCGGATGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
758 AGAAGATCAAGCGCGCTGACCGCGCATCTGCGAGGAGATGGAAGAGGCGCGAGATCACCA 817  
601 AGAAGATCAAGCGCGCTGATGAGATCTGCAACCGGAGTGGAGAGGCGCGAGATCACCA 660  
818 AGATCGCGCGCGGAGACCGCTTACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877  
661 AGATCGCGCGCGGAGACCGCTTACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720  
878 CCAAGTGGCGCAAGCTGTGTGAATTCGCGAGCTGGAACAAGCGCGCACCGAGCTTCTGGG 937  
721 CCAAGTGGCGCAAGCTGTGTGAATTCGCGAGCTGGAACAAGCGCGCACCGAGCTTCTGGG 780  
938 AGGTGAGCTGGGCGATCCCG 997  
781 AGGTGAGCTGGGCGATCCCG 840  
998 TGGAGCTGGGCGCGCTTACTTCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057  
841 TGGAGCTGGGCGCGCTTACTTCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
1058 CTTTACCATCCCGAGCATCAACAAAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1117  
901 CTTTACCATCCCGAGCATCAACAAAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
1118 TGCCCCAGGCTGGAAGGCG 1177  
961 TGCCCCAGGCTGGAAGGCG 1020

```
Qy 1178 AGCCCTTCGCGCCGCGCAACCCCGAGATCGTGATCTACCA-----GCCCCCTGTACG 1231
Db 1021 AGCCCTTCGCGCAAGCAGAACCCCGACATCGTGTCTACAGTACATGACGACCTGTACG 1080
Qy 1232 TGGGCAAGGACCTGGAGATCGGCCAGCACCGCCCAAGATCAGGAGCTGGCAGACACC 1291
Db 1081 TGGGCAAGGACCTGGAGATCGGCCAGCACCGCCCAAGATCAGGAGCTGGCAGACACC 1140
Qy 1292 TGCTGCGCTGGGGCTTCAACACCCCGCAAGAAGCACAGAGAGAGCCCCCTTCCTGT 1351
Db 1141 TGCTGCGCTGGGGCTTCAACACCCCGCAAGAAGCACAGAGAGAGCCCCCTTCCTGT 1200
Qy 1352 GGATGGGCTACGAGCTGACCCCGGACCAAGTGGAACGCTGACGCCATCAGCTGCCCGAGA 1411
Db 1201 GGATGGGCTACGAGCTGACCCCGGACCAAGTGGAACGCTGACGCCATCAGCTGCCCGAGA 1260
Qy 1412 AGGAGAGCTGACCGTGAACGACATCCAGAGAGCTGGTGGCAAGCTGAACCTGGCCAGCC 1471
Db 1261 AGGAGAGCTGACCGTGAACGACATCCAGAGAGCTGGTGGCAAGCTGAACCTGGCCAGCC 1320
Qy 1472 AGATCTACCCCGGATCAAGGTGGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCC 1531
Db 1321 AGATCTACCCCGGATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCC 1380
Qy 1532 TGACCGACATCGTGCCCTGACCGGAGGCGGAGCTGGAGCTGGCCGAGAACCGCGAGA 1591
Db 1381 TGACCGAGGTGATCCCTGACCGGAGGCGGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440
Qy 1592 TCCTGCGGAGCGGTGACCGGCTGTACTAGACCCCGACCAAGGACCTGGTGCCCGAGA 1651
Db 1441 TCCTGAAAGGAGCCGTGACAGAGTGTACTAGACCCCGACCAAGGACCTGGTGCCCGAGA 1500
Qy 1652 TCCAGAGCAGGGCCACGACAGTGGACTTACAGATCTACAGAGAGCCCTTCAAGAAC 1711
Db 1501 TCCAGAGCAGGGCCACGACAGTGGACTTACAGATCTACAGAGAGCCCTTCAAGAAC 1560
Qy 1712 TGAAGACGGGCAAGTACGCCAAGATGGCGACCGGCCACCAACGAGCGTGAAGCAGCTGA 1771
Db 1561 TGAAGACGGGCAAGTACGCCGAGTACGCCGCGGCCACCAACGAGCGTGAAGCAGCTGA 1620
Qy 1772 CCGAGGCGGTGCAGAGATCCCATGGAGAGCATGTGTATCTGGGGCAGACCCCAAGT 1831
Db 1621 CCGAGGCGGTGCAGAGGTGAGCACCGGAGAGCATGTGTATCTGGGGCAGAGTATCCCAAGT 1680
Qy 1832 TCCGCTGCCCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTACTTGGCAGGCCA 1891
Db 1681 TCAAGCTGCCCATCCAGAGGAGACCTGGGAGGCTGTGTGATGGAGTACTTGGCAGGCCA 1740
Qy 1892 CCTGGATCCCCGAGTGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGTACCAAGC 1951
Db 1741 CCTGGATCCCCGAGTGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGTACCAAGC 1800
Qy 1952 TGGAGAGGAGCCCATCATCGGCGCGAGACTTCTAGTGAAGCGGCGCCGCAACCGCG 2011
Db 1801 TGGAGAGGAGCCCATCATCGTGGCGCGAGACTTCTAGTGAAGCGGCGCCGCAACCGCG 1860
Qy 2012 AGACCAAGATCGGCAAGCGCGGTACGTGACCGACCGGCGCGGCGAGAGATCGTGAGCC 2071
Db 1861 AGACCAAGCTGGGCAAGCGCGGTACGTGACCGACCGGCGCGGCGAGAGGTGTGAGCA 1920
Qy 2072 TGACCCGAGACCAACCAAGAGACCGGAGCTGCGAGGCCATCCAGCTGGCCCTGAGGACA 2131
Db 1921 TCGCCGACACCAACCAAGAGACCGGAGCTGCGAGGCCATCCAGCTGGCCCTGAGGACA 1980
Qy 2132 CGCGAGGAGGTGAACATCTGTGACCGACAGCAGTACGCCCTGGGCAATCATCAGGCC 2191
Db 1981 CGCGCCTGGAGGTGAACATCTGTGACCGACAGCAGTACGCCCTGGGCAATCATCAGGCC 2040
Qy 2192 AGCCCGACAGAGCGAGCGAGCTGTGTGAACACAGATCATCAGCAGCTGATCAGAGG 2251
Db 2041 AGCCCGACAGAGCGAGCGAGCTGTGTGAACACAGATCATCAGCAGCTGATCAGAGG 2100
```

```
Qy 2252 AGAAGGTGTACTGAGCTGGTGCCTCCGCCCAAGAGGATCGCGGGCAACGAGCAGATCG 2311
Db 2101 AGAAGGTGTACTGAGCTGGTGCCTCCGCCCAAGAGGATCGCGGGCAACGAGCAGGTGG 2160
Qy 2312 ACAAGCTGGTGAAGCAAGGATCGCAAGGTGCTTCTCGAAGGATCGATGCGCGCA 2371
Db 2161 ACAAGCTGGTGAAGCGCGGATCGCAAGGTGCTTCTCGAAGGATCGATGCGCGCA 2220
Qy 2372 TCCTGATCTACAGTATACGACCACTGTGAGTGGGAGCGCGGCTTAGGATCGATT 2431
Db 2221 TCCTGATCTACAGTATACGACCACTGTGAGTGGGAGCGCGGCTTAGGATCGATT 2280
Qy 2432 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2463
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 10
AR373389
LOCUS AR373389 2312 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 84 from patent US 6602705.
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like
particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
FEATURES
source
1. .2312
/organism="unknown"
/moi_type="genomic DNA"

Query Match 82.2%; Score 2025.2; DB 6; Length 2312;
Best Local Similarity 93.3%; Pred. No. 1.7e-202;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

Qy 170 GCGCAAGGAGGGCCACAGATGAAGACTGCCAGGCGCCAGGCCAACTTCTTCCGCG 229
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAAATTTCTTCCGCG 60
Qy 230 AGGACCTGGCTTCCCGCAGGCAAGCCCGAGTTCCCGAGGAGCAGAACCGCGCA 289
Db 61 AGGACCTGGCTTCTCGAGGCAAGCCCGAGTTTCCAGGCGAGCGAGCGACCGCGCA 120
Qy 290 ACAGCCCCACAGCCGCGAGCTGCAGGTGCGCGCG-----ACAACCCCCGACGAGG 343
Db 121 ACAGCCCCACCGCGCGAGCTGCAGGTGTTGGGCGCGGAGAACAAACAGCTGAGCGAGG 180
Qy 344 CCGGCGCGAGCGCCAGGCGACCCCTG-----AACTTCCCCAGATCACTCTGTGGCAGC 397
Db 181 CCGGCGCGAGCGCCAGGCGACCGCTGAGCTTCAACTTCCCCAGATCACTCTGTGGCAGC 240
Qy 398 GCGCCCTGTGTAGCATCAAGTGGCGCGCAGATCAAGGAGGCGCTGCTGCACACCGCGC 457
Db 241 GCGCCCTGTGTAGCATCAAGTGGCGCGCAGCTCAAGGAGGCGCTGCTGCACACCGCGC 300
Qy 458 CCGACGACACCGTCTCGAGGAGATGAGCTTCCCGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGACGACACCGTCTCGAGGAGATGAACCTGCGCGCAAGTGAAGCCCAAGATGATCG 360
Qy 518 GCGGATCGGCGGCTTCAAGGTGCGCAGTACGACAGATCCTGATCGAGATCTGCG 577
Db 361 GCGGATCGGCGGCTTCAAGGTGCGCAGTACGACAGATCCTGATCGGAGATCTGCG 420
Qy 578 GCAAGAGGCGCATCGGCGACCGTGTGATCGGCGCGCCCGCTGAACATCATCGGCGCA 637
Db 421 GCCAAGGCGCATCGGCGACCGTGTGTTGGGCGCCCAACCCCGGTGAACATCATCGGCGCA 480
```



Qy 638 ACATGCTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCCCATCAGCCCCCATCGAGACCGTGC 697  
Db 481 ACCTGCTGACCCAGATCGGCTGCACCCCTGAACCTTCCCCCATCAGCCCCCATCGAGACCGTGC 540  
Qy 698 CCGTGAAGCTGAAGCCCGGCGATGGACGGCCCGCCCAAGGTGAAGCTGAGTGGCCCTCGACGAGG 757  
Db 541 CCGTGAAGCTGAAGCCCGGCGATGGACGGCCCGCCCAAGGTCAAGAGTCAAGCAGTGGCCCTCGACGAGG 600  
Qy 758 AGAAGATCAAGGCGCTGACCGCCATCTGCAGGAGATGGAGAGGAGGGCAAGATCACCA 817  
Db 601 AGAAGATCAAGGCGCTGCTGGAGATCTGCACCGAGATGGAGAGGAGGGCAAGATCACGA 660  
Qy 818 AGATCGGCCCGGAGAACCCCTACAACACCCCGCTGTTCCGCCATCAAGAAGAAGACAGCA 877  
Db 661 AGATCGGCCCGGAGAACCCCTACACACCCCGCTGTTCCGCCATCAAGAAGAAGACAGCA 720  
Qy 878 CCAAGTGGCGCAAGCTGGTGGACTTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 937  
Db 721 CCAAGTGGCGCAAGCTGGTGGACTTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 780  
Qy 938 AGTGCAGCTGGGCAATCCCCGACCCCGCGGCTGAAAGAAAGAGAGCGTGAACCGTGC 997  
Db 781 AGTGCAGCTGGGCAATCCCCGACCCCGCGGCTGAAAGAAAGAGAGCGTGAACCGTGC 840  
Qy 998 TGGACGTGGGCGAGCGCTACTTCAAGCGTGCCTCGGACGAGGACTTCCGCAAGTACACCG 1057  
Db 841 TGGACGTGGGCGAGCGCTACTTCAAGCGTGCCTCGGACGAGGACTTCCGCAAGTACACCG 900  
Qy 1058 CCTTACCATCCCGAGCATCAACAACGAGACCCCGGCAATCCGCTACAGTACAAAGTGC 1117  
Db 901 CCTTACCATCCCGAGCATCAACAACGAGACCCCGGCAATCCGCTACAGTACAAAGTGC 960  
Qy 1118 TGCCCCAGGCTGGAAGGCGAGCCCCAGATCTTCCAGAGCAGCATCAACAGATCTGCG 1177  
Db 961 TGCCCCAGGCTGGAAGGCGAGCCCCAGATCTTCCAGAGCAGCATCAACAGATCTGCG 1020  
Qy 1178 AGCCCTTCCGCGCCGCAACCCCGAGATCTGTACTACCA-----GGCCCGCCCTGTACG 1231  
Db 1021 AGCCCTTCCGCGAAGCAACCCCGACATCTGTACTACCAAGTACATGGAAGCACTGTACG 1080  
Qy 1232 TGGCAGCGACCTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTCGCAAGCAC 1291  
Db 1081 TGGCAGCGACCTGGAGATCGGCGAGCACCGCAAGATCGAGGAGCTCGCGCAGCAC 1140  
Qy 1292 TGCTGCGCTGGGCTTCAACACCCCGCAAGAAGACACAGAGGAGCCCCCTTCTGT 1351  
Db 1141 TGCTGCGCTGGGCTTCAACACCCCGCAAGAAGACACAGAGGAGCCCCCTTCTGT 1200  
Qy 1352 GGATGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGA 1411  
Db 1201 GGATGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGA 1260  
Qy 1412 AGGAGAGCTGGAACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAATGGGCGCACGC 1471  
Db 1261 AGGACAGCTGGAACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAATGGGCGCACGC 1320  
Qy 1472 AGATCTACCCCGGATCAAGGTGGCGGCGAGCTGTGCAAGCTGTGCGCGGCGCCCAAGGCC 1531  
Db 1321 AGATCTACCCCGGATCAAGGTGGAAGCAGCTGTGCAAGCTGTGCGCGGCGCCCAAGGCC 1380  
Qy 1532 TGACCGACATCGTCCCTGCACGAGGAGCGGAGCTGGAGCTGGCGCGGAGAACCGCGAGA 1591  
Db 1381 TGACCGAGGTGATCCCTCTGACCGAGGAGCGGAGCTGGAGCTGGCGCGGAGAACCGCGAGA 1440  
Qy 1592 TCCTGCGGAGCGCGTGCAGCGGCTGTACTACGACCCCGACAGGAGCTGTGGTGGCGGAGA 1651  
Db 1441 TCCTGAAGGAGCGCGTGCAGGCTGTACTACGACCCCGACAGGAGCTGTGGTGGCGGAGA 1500  
Qy 1652 TCCGAAGCAGGCGCACGACGAGTGGACCTTACCGAGTCTACGAGGAGCCCTTCCAGAAC 1711  
Db 1501 TCCGAAGCAGGCGCACGAGTGGAGCTTACCGAGTCTACGAGGAGCCCTTCCAGAAC 1560

Qy 1712 TGAAGACCGGCAAGTAGTACGCCAAGATGCGCACCGCCCCACACCAACGACGTCGAGCAGCTGA 1771  
Db 1561 TGAAGACCGGCAAGTAGTACGCCCGCATGCGGGCGCCCCACACCAACGACGTCGAGCAGCTGA 1620  
Qy 1772 CCAGAGCGCTGCAAGATCGCCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCCCAAGT 1831  
Db 1621 CCAGAGCGCTGCAAGAGGTGAGCACCGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGT 1680  
Qy 1832 TCCGCTCGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGAGCCGACTACTTGGCAGGCCA 1891  
Db 1681 TCAAGCTTGCCCATCCAGAAGGAGACCTGGGAGGCGCTGGTGATGGAGTACTTGGCAGGCCA 1740  
Qy 1892 CTTGGATCCCCCGAGTGGGAGTCTGTGAACAACCCCCCCCCCTGGTGAAGCTGTGGTACACG 1951  
Db 1741 CTTGGATCCCCCGAGTGGGAGTCTGTGAACAACCCCCCCCCCTGGTGAAGCTGTGGTACACG 1800  
Qy 1952 TGAGAAGGAGCCATCATCGGCGCGAGACCTTCTAAGTGGAGCGGCGCCGCAACCGCG 2011  
Db 1801 TGAGAAGGAGCCATCATCGTGGCGCGAGACCTTCTAAGTGGAGCGGCGCCGCAACCGCG 1860  
Qy 2012 AGACCAAGATCGGCAAGCGGCTAGTGACCGACCGGGCGCGGACGAAGATCGTGAGCC 2071  
Db 1861 AGACCAAGCTGGGCAAGCGGCTAGTGACCGACCGGGCGCGGACGAAGTGGTGAGCA 1920  
Qy 2072 TGACCGAGACCAACAACGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGGACA 2131  
Db 1921 TCGCCGACACCAACAACGAGACCGAGCTGCAGGCGCATCCACCTGGCCCTGCAGGACA 1980  
Qy 2132 GCGCAGCGAGGTGAACATCGTGACCGAGCAGCAGTACGCCCTGGGGCATATCCAGGCC 2191  
Db 1981 GCGCCTTGGAGGTGAACATCGTGACCGAGCAGCAGTACGCCCTGGGGCATATCCAGGCC 2040  
Qy 2192 AGCCGACAGAGCGAGAGCGAGCTGTGAACCAAGATCATCGAGCAGCTCATCAAGAAGG 2251  
Db 2041 AGCCGACAGAGCGAGAGCGAGCTGTGAGCCAGATCATCGAGCAGCTCATCAAGAAGG 2100  
Qy 2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGCGCATCGGCGGCAACGAGCAGATCG 2311  
Db 2101 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGCGCATCGGCGGCAACGAGCAGGTGG 2160  
Qy 2312 ACAAGCTGTGAGCAAGGGCATCCGCAAGTGTCTTCTGGAGCGGATCGATGGCGGCA 2371  
Db 2161 ACAAGCTGTGAGCGCGGATCCGCAAGGTGTCTTCTGAAACGGCATCGATGGCGGCA 2220  
Qy 2372 TCGTGTATCTACAGTACATGAGCAGCTGTACGTGGGCGGCGCTAGGATCGATT 2431  
Db 2221 TCGTGTATCTACAGTACATGAGCAGCTGTACGTGGGCGGCGCTAGGATCGATT 2280  
Qy 2432 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2463  
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 11  
BD263705

LOCUS  
DEFINITION

BD263705 2300 bp DNA linear PAT 17-JUL-2003  
Improved expression of HIV polypeptides and production of  
virus-like particles.

ACCESSION  
VERSION

BD263705  
BD263705.1 GI:33073473  
JP 2002533124-A/72.

KEYWORDS  
SOURCE

synthetic construct  
other sequences; artificial sequences.

ORGANISM

REFERENCE  
AUTHORS

1 (bases 1 to 2300)  
Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,  
Greer,C., Selby,M. and Walker,C.

TITLE

Improved expression of HIV polypeptides and production of  
virus-like particles

JOURNAL

Patent: JP 2002533124-A 72 08-OCT-2002;  
CHIRON CORP  
OS Artificial Sequence

COMMENT

PN JP 2002533124-A/72

PD	08-OCT-2002
PF	30-DEC-1998 JP 2000591193
PR	31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI SUSAN BARNETT, JAN ZUR WEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI KARIN HARTOG,
PI	HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC C12N5/09,A61K13/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC C12N5/10,
PC	C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC Description of Artificial Sequence: PS(-).protmod.RTopt.YMM FH
Key	Location/Qualifiers
FT	source 1..2300 /organism='Artificial Sequence'.
FEATURES	Location/Qualifiers
source	1..2300 /organism="synthetic construct" /mol_type="genomic DNA" /db_xref="taxon:32630"
ORIGIN	
Query Match	82.0%; Score 2019.2; DB 6; Length 2300;
Best Local Similarity	93.2%; Pred. No. 7,1e-202;
Matches 2150;	Conservative 0; Mismatches 138; Indels 18; Gaps 3
Qy	170 GCGGCAAGGAGGGCCACAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGG 229
Db	
Qy	1 GCGGCGCGAAGGACACCAAATGAAGAATTGCACCTGAGAGACAGGCTAATTTCTTCGCGG 60
Db	
Qy	230 AGGACCTGGCCCTCCCCAGAGGCAAGGCCGCGAGTTCCCGACGAGCAGACACCGGCCA 289
Db	
Qy	61 AGGACCTGGCCCTCTCTGAGGGCAAGGCCGCGAGTTTCAGCAGCAGCAGACCCCAGGCCA 120
Db	
Qy	290 ACAGCCCCACCAGCCGCGAGCTGCAGGTGCGCGCG-----ACAACCCCCGACGCGAGG 343
Db	
Qy	121 ACAGCCCCACC CGCGCGAGCTGCAGGTGCGGCGCGAGAACACAGCCTGACGCGAGG 180
Db	
Qy	344 CCGGCGCGAGCGCCAGGGCACCCCTG-----AACTTCCCCAGATCACTCTGTGGGACG 397
Db	
Qy	181 CCGGCGCGCGACCGCGCAGGGCACCGTGAGCTTCAACTTCCCCAGATCACTCTGTGGCAGC 240
Db	
Qy	398 GCCCCTGTGTAGCATCAAGTGTGGCGGCGCAGATCAAGAGGCGCCTGTGGACACCGGCG 457
Db	
Qy	241 GCCCCTGTGTGACCATCAGATCGCGCGCCAGCTCAAGAGGCGCTGTCTCGACACCGGCG 300
Db	
Qy	458 CCGACGACACCGTCTGTGAGGAGATGAGCCTGCCCGCAAGTGGAGGCCAAGATGATCG 517
Db	
Qy	301 CCGACGACACCGTCTGTGAGGAGATGAACCTGCCCGCGCAGTGGGAAGCCAAGATGATCG 360
Db	
Qy	518 GCGGCATCGCGCGCTTCATCAAGGTGCGGCCAGTAGCACAGATCTCTGATCGAGATCTCGG 577
Db	
Qy	361 GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTAGCACAGATCCCGCTGGAGATCTGCG 420
Db	
Qy	578 GCAGAAGGCCATCGGCACCGTGTGATCGGGCCCCACCCCGTGAAACATCATCGGGCCGCA 637
Db	
Qy	421 GCCACAAAGGCCATCGGCACCGTGTGTGGGGCCCCACCCCGTGAAACATCATCGGCCCGCA 480
Db	
Qy	638 ACATGTGACCCAGCTGGGGCTGCACCCCTGAACCTTCCCCCATCAGCCCCCATCGAGACCGTGC 697
Db	
Qy	481 ACTTGTGTACCCAGATCGGCTGTGACCTTCCCATCAGGCCCATCGAGACCGGTGC 540
Db	
Qy	698 CCGTGAAGCTGAAGCCCGGCATGACGCGCCCCAAGGTGAAGCAGTGGGCCCCCTGACCCGAGG 757
Db	
Qy	541 CCGTGAAGCTGAAGCCGGGATGGACGCGCCCCAAGGTCAAGCAGTGGGCCCCCTGACCCGAGG 600
Db	
Qy	758 AGAGATCAAGGCCCTGACCGCCATCTCTGAGGAGATGGAGAGGAGGGCAAGATCAACA 817
Db	
Qy	601 AGAAGATCAAGGCCCTGTGTGAGATCTGTGCAACCGAGATGGAGAGAGGGCAAGATCAACA 660
Db	
Qy	818 AGATCGGCCCGGAGAACCCCTTACACACCCCGCTGTTCGCCATCAAGAAAGAGGACAGCA 877
Db	
Qy	661 AGATCGGCCCGGAGNAACCCCTTACACACCCCGCTGTTCGCCATCAAGAAAGAGGACAGCA 720
Db	
Qy	878 CCAAGTGGCGCAAGCTGGTGGACTTTCGCGGAGCTTGAAACAAGCGCACCCAGGACTTCTCGG 937

Db 1795 AGGAGCCCATCTGTGGCGCCGAGACCTTCTACGTGACGCGCGCCCAACCGGAGACCA 1854  
Qy 2018 AGATTGGCAAGCGCGCTAGTACACCGACCGGGCGGAGAGATCGTGAAGCTGACCG 2077  
Db 1855 AGCTGGGCAAGCGCGCTAGTACACCGACCGGGCGGAGAGTGGTGAAGATCGCCG 1914  
Qy 2078 AGACCACCAACCAAGAACCGAGCTGACGAGCCATCCACAGCTGGCCCTGCAGGACAGCGGCA 2137  
Db 1915 ACACCACCAACCAAGAACCGAGCTGACGAGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974  
Qy 2138 GCGAGGTGAACATCTGTACCGACAGCCAGCAGTACGCCCTTGGGGATCATCCAGGCCAGCCCG 2197  
Db 1975 TGGAGGTGAACATCTGTACCGACAGCCAGCAGTACGCCCTTGGGCATCATCCAGGCCAGCCCG 2034  
Qy 2198 ACAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2257  
Db 2035 ACAGAGCGAGAGCGAGCTGGTGAAGCAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2094  
Qy 2258 TGTACCTGAGCTGGGTGCCCGCCCAAGAGGCGATCGCGGCAACGAGCAGATCGACAAGC 2317  
Db 2095 TGTACCTGGCTGGGTGCCCGCCCAAGAGGCGATCGCGGCAACGAGCAGTGAACAAGC 2154  
Qy 2318 TGTGAGCAAGGGGATCCGCAAGGTGCTGTCTCTGGAACGGGATCGATGGCGGATCGTGA 2377  
Db 2155 TGTGAGCGCGGCATCCGCAAGGTGCTGTCTCTGGAACGGGATCGATGGCGGATCGTGA 2214  
Qy 2378 TCTACAGTACATGAGCAGCTGTACGTGGGCGAGCGCGGCCCTAGGATCGATTAAAGC 2437  
Db 2215 TCTACAGTACATGAGCAGCTGTACGTGGGCGAGCGCGGCCCTAGGATCGATTAAAGC 2274  
Qy 2438 TTCCCGGGCTAGCACCGGTGAATTC 2463  
Db 2275 TTCCCGGGCTAGCACCGGTGAATTC 2300

## RESULT 12

CQ870575  
LOCUS CQ870575 2300 bp DNA linear PAT 13-SEP-2004  
DEFINITION Sequence 83 from Patent EP1433851.  
ACCESSION CQ870575  
VERSION CQ870575.1 GI:52000091  
KEYWORDS synthetic construct  
SOURCE other sequences; artificial sequences.  
ORGANISM  
REFERENCE 1  
AUTHORS Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,  
Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: EP 1433851-A 83 30-JUN-2004;  
CHIRON CORPORATION (US)  
FEATURES  
source  
1. 2300  
/location/Qualifiers  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Description of Artificial Sequence:  
FS(-).protmod.Rtopt.YMMW"

## ORIGIN

Query Match 82.0%; Score 2019.2; DB 6; Length 2300;  
Best Local Similarity 93.2%; Pred. No. 7.1e-202;  
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
Qy 170 GCGCAAGAGGCGCCACCATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCGCG 229  
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGCAGGCTAAATTTCTTCGCG 60  
Qy 230 AGGACCTGGCTTCCCCCAGGGCAAGCGCCGCGAGTTCCCCAGGAGCAGAACCGCGCCA 289  
Db 61 AGGACCTGGCTTCTCTCAGGCGCAAGCGCCGCGAGTTTCAGCAGCGAGCAGACCGCGCCA 120

Qy 290 ACAGCCCCACACGACCCCGAGCTCGAGTGGCGCGCGG-----ACAAACCCCGCAGCGAGG 343  
Db 121 ACAGCCCCACACCGCGCGAGCTCGAGTGGCGCGCGGAGAAACAAACAGCCTGAGCGAGG 180  
Qy 344 CCGGCGCGAGCGCGAGGGACCGCTG-----NACTTCCCCCAGATCACCTGTGGCAGC 397  
Db 181 CCGGCGCGAGCGCGAGGGCACCGTTCAGCTTCACTTCCCCCAGATCACCTGTGGCAGC 240  
Qy 398 GCCCCTGTGTGACATCAAGGTGGCGCGCGAGATCAAGGAGGCGCTGTCTGACACCGCGC 457  
Db 241 GCCCCTGTGTGACATCAAGGTGGCGCGCGAGTACGACAGATCTCTGATCGAGATCTGCG 300  
Qy 458 CCGACACACCGCTGTCTGGAGGAGATGAGCCTGCGCGCAAGTGGAAAGCCCAAGATGATCG 517  
Db 301 CCGACGACACCGTGTGGAGGAGATGAACCTGCGCGCAAGTGGAAAGCCCAAGATGATCG 360  
Qy 518 GCGGCATCGCGCGCTTCATCAAGGTGGCGCGCGAGTACGACAGATCTCTGATCGAGATCTGCG 577  
Db 361 GCGGCATCGCGCGCTTCATCAAGGTGGCGCGCGAGTACGACAGATCTCTGATCGAGATCTGCG 420  
Qy 578 GCAAGAGGCGCATCGGCACCGCTGATCGCGCGCCACCCCGGTGAACATCATCGGCGCA 637  
Db 421 GCCAAGAGGCGCATCGGCACCGCTGATCGCGCGCCACCCCGGTGAACATCATCGGCGCA 480  
Qy 638 ACATGTCTGACCCAGCTGGGCTGCACCTGAACTTTCCCATCAGCCCCATCGAGACCGTGC 697  
Db 481 ACCTGTCTGACCCAGCTGGGCTGCACCTGAACTTTCCCATCAGCCCCATCGAGACCGTGC 540  
Qy 698 CCGTGAAGCTGAAGCCCGGATGAGCGCGCCCAAGGTCAAGCAGTGGCCCCCTGACCGAGG 757  
Db 541 CCGTGAAGCTGAAGCCCGGATGAGCGCGCCCAAGGTCAAGCAGTGGCCCCCTGACCGAGG 600  
Qy 758 AGAAGATCAAGGCCCTGACCGCATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCA 817  
Db 601 AGAAGATCAAGGCCCTGCGTGGAGATCTGCAACGAGATGGAGAGGAGGCGCAAGATCACCA 660  
Qy 818 AGATCGGCCCCGAGAACCCCTCAACAACCCCGTGTTCGCCATCAAGAAGAAGACAGCA 877  
Db 661 AGATCGGCCCCGAGAACCCCTCAACAACCCCGTGTTCGCCATCAAGAAGAAGACAGCA 720  
Qy 878 CCAAGTGGCGCAAGCTGGTGGACTTCGGCGAGCTGAACAAGCGCACCCAGACTTCTGGG 937  
Db 721 CCAAGTGGCGCAAGCTGGTGGACTTCGGCGAGCTGAACAAGCGCACCCAGACTTCTGGG 780  
Qy 938 AGGTGCAAGCTGGGCATCCCCCACCGCGCTGAAGAAGAAGAAGAGCGTCAACCGTGC 997  
Db 781 AGGTGCAAGCTGGGCATCCCCCACCGCGCTGAAGAAGAAGAAGAGAGTGAACCGTGC 840  
Qy 998 TGGACGTGGCGCAGCGCTACTTCAAGGTGCGCTTGGACGAGGACTTCCGCAAGTACACCG 1057  
Db 841 TGGACGTGGCGCAGCGCTACTTCAAGGTGCGCTTGGACGAGGACTTCCGCAAGTACACCG 900  
Qy 1058 CTTTACCATCCCCCAGATCAACAACGAGACCCCGGATCCCGTACCACTACCACTGC 1117  
Db 901 CTTTACCATCCCCCAGATCAACAACGAGACCCCGGATCCCGTACCACTACCACTGC 960  
Qy 1118 TGCCCGAGGCGTGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATCAACAAGATCTCTGG 1177  
Db 961 TGCCCGAGGCGTGAAGGGCAGCCCCGCGCATCTTCCAGAGCAGCATCAACAAGATCTCTGG 1020  
Qy 1178 AGCCCTTCCGCGCCGCAACCCCGAGATCTGATCTACACAGGCCCCCTCTGTAGTGGGCA 1237  
Db 1021 AGCCCTTCCGCAAGCAGACAGACCCCGCATCTGATCTACACAGGCCCCCTCTGTAGTGGGCA 1080  
Qy 1238 GCGACCTGGAGATCGGCGCAGCAGCGCGCAAGATCGAGGAGCTGGCGAAGCAGCTGCTGC 1297  
Db 1081 GCGACCTGGAGATCGGCGCAGCAGCGCAACCAAGATCGAGGAGCTGGCGAAGCAGCTGCTGC 1140  
Qy 1298 GCTTGGGCTTTCACACCCCGCAAGAGCAACAGAGAGGAGGCCCCCTTCTCTGTGATGG 1357  
Db 1141 GCTTGGGCTTTCACACCCCGCAAGAGCAACAGAGAGGAGGCCCCCTTCTCTGTGATGG 1198  
Qy 1358 GCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGGCCATCGAGCTGCCCCCGAGAGAGGA 1417

Db 1199 -----CGAGCTGCACCCGACAAAGTGGACCGTGCAGCCCATCATGCTGCGCCGAGGAAGACA 1254  
 Qy 1418 GCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477  
 Db 1255 GCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1314  
 Qy 1478 ACCCGGGGATCAAGGTGGCCGAGCTGTGCAAGCTGTGCGGGGGCCCAAGGCCCTGACCG 1537  
 Db 1315 ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGGGGCAACCAAGGCCCTGACCG 1374  
 Qy 1538 ACATCGTCCCTGACCGAGGAGCGCAGCTGGAGCTGGCCGAGAAACCGGAGATCCTGC 1597  
 Db 1375 AGTGATCCCCCTGACCGAGAGCGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA 1434  
 Qy 1598 GCGAGCCGCTGACCGGCTGTACTACGACCCAGCAAGGACTGTGTGGCCGAGATCCAGA 1657  
 Db 1435 AGGAGCCCGTGACGAGGTGTACTACGACCCAGCAAGGACTGTGTGGCCGAGATCCAGA 1494  
 Qy 1658 AGCAGGGCCAGCAGTGGACCTTACGAGATCTTACGAGGAGCCCTTCAAGAACTGGA 1717  
 Db 1495 AGCAGGGCCAGGGCCAGTGGACCTTACGAGATCTTACGAGGAGCCCTTCAAGAACTGGA 1554  
 Qy 1718 CCGCAAGTACGCAAGTGGACCGCACCGCCACACCAAGCAGCTGAAGCTGACCGAG 1777  
 Db 1555 CCGCAAGTACGCGCCGATGCGCGCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGG 1614  
 Qy 1778 CCGTGCAAGAGATCGCATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCC 1837  
 Db 1615 CCGTGCAAGAGTGAACACCGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674  
 Qy 1838 TGCCCATCCAGAAAGGAGACTGGGAGACTGTGTGGACCGACTACTGGCAGGCCACCTGGA 1897  
 Db 1675 TGCCCATCCAGAAAGGAGACTGGGAGCTGTGTGATGTGAGTACTTGGCAGGCCACCTGGA 1734  
 Qy 1898 TCCCGGAGTGGAGTTGTGAAACACCCGCCCTTGTGAGCTGTGTGAGCTGTGTGAGTGA 1957  
 Db 1735 TCCCGGAGTGGAGTTGTGAAACACCCGCCCTTGTGAGCTGTGTGAGTGTGAGTGA 1794  
 Qy 1958 AGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCA 2017  
 Db 1795 AGGAGCCCATCTGTGGGCGCGAGACTTCTACGTGGAGCGCGCCCAACCGCGAGACCA 1854  
 Qy 2018 AGATCGCAAGCGCGCTAGTGACCAAGCGGCGCCGAGAGATCGTGAGCTGTGACCG 2077  
 Db 1855 AGCTGGGCAAGCGCGCTAGTGACCAAGCGGCGCCGAGAGCTGTGTGAGCATCGCG 1914  
 Qy 2078 AGACCAACCAAGACCGAGCTGACGCGCATCCAGCTGGCCCTTGCAGGACAGCGGCA 2137  
 Db 1915 ACACCAACCAAGACCGAGCTGACGCGCATCCACCTGGCCCTTGCAGGACAGCGGCG 1974  
 Qy 2138 GCGAGGTGAACATCGTGACCGACAGCAGTACGCGCTTGGGCTATCCAGGCCAGCGCG 2197  
 Db 1975 TGGAGGTGAACATCGTGACCGACAGCAGTACGCGCTTGGGCTATCCAGGCCAGCGCG 2034  
 Qy 2198 ACAAGCGAGGAGCGAGCTGTGTGAACCAAGCAGTATCGAGCAGCTGATCAAGAGGAGAGG 2257  
 Db 2035 ACAAGCGAGGAGCGAGCTGTGTGAACCAAGCAGTATCGAGCAGCTGATCAAGAGGAGAGG 2094  
 Qy 2258 TGTAACCTGAGTGGTGGCGCCCAAGGCGATCGGGGGCAACGAGCAGATCGCAAGC 2317  
 Db 2095 TGTAACCTGAGTGGTGGCGCCCAAGGCGATCGGGGGCAACGAGCAGTGGTGAACG 2154  
 Qy 2318 TGGTGAACAGGGGATCCGCAAGGTGTGTCTTGGACGGCATCGATGATGGGGCATCGTGA 2377  
 Db 2155 TGGTGAAGCGCGGATCCGCAAGGTGTGTCTTGGACGGCATCGATGATGGGGCATCGTGA 2214  
 Qy 2378 TCTACAGTATACGAGCAGCTGTACGTGGGACGCGCGCCCTAGGATCGATTAAAGC 2437  
 Db 2215 TCTACAGTATACGAGCAGCTGTACGTGGGACGCGCGCGCCCTAGGATCGATTAAAGC 2274  
 Qy 2438 TTCCCGGGGTAGCACCGGTGAATTC 2463  
 |||||

Db 2275 TTCCCGGGGTAGCACCGGTGAATTC 2300  
 RESULT 13  
 AR373388  
 LOCUS AR373388 2300 bp DNA linear PAT 18-DEC-2003  
 DEFINITION Sequence 83 from patent US 6602705.  
 ACCESSION AR373388  
 VERSION AR373388.1 GI:40075491  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 2300)  
 AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.  
 TITLE Expression of HIV polypeptides and production of virus-like particles  
 JOURNAL Patent: US 6602705-A 83 05-AUG-2003;  
 FEATURES  
 source Location/Qualifiers  
 1..2300  
 /organism="unknown"  
 /mol\_type="genomic DNA"  
 ORIGIN  
 Query Match 82.0%; Score 2019.2; DB 6; Length 2300;  
 Best Local Similarity 93.2%; Pred. No. 7,1e-202;  
 Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
 Qy 170 GCGCAAGGAGGGCCACAGATGAAGACTGCACCGAGCGCCAGCCAACTTCTTCGCG 229  
 Db 1 GCGGCCCGGAAGGACCAATGAAGATTGCATGAGAGACAGGTAATTTCTTCGCG 60  
 Qy 230 AGGACCTGGCTTCTCCCGCAGGGCAAGCCCGCAGTTTCCCGAGCGCAGCAACCGCGCA 289  
 Db 61 AGNACCTGGCTTCTTCAGGGCAAGCCCGCAGTTTTCAGCAGCAGCAGACCCGCGCA 120  
 Qy 290 ACAGCCCCACAGCCGCGAGCTGCAGTGCAGCGCGGCGGCGGCGGCGGCGGCGG 343  
 Db 121 ACAGCCCCACCGCGCGAGCTGCAGTGTGGGCGCGGAGAACACAGCCTGAGCGAGG 180  
 Qy 344 CCGCGCGCGAGCGCCAGGGCACCTCGGCGGCGCGCGCGCGCGCGCGCGCGCGCG 397  
 Db 181 CCGCGCGCGAGCGCCAGGGCACCTGAGCTTCAATTTCCCGCAGATCACCTGTGGCAGC 240  
 Qy 398 GCCCTCTGTGTAGCATCAAGGTGGCGCGCCAGATCAAGAGGCGCTGTGTGACACCGCG 457  
 Db 241 GCCCTCTGTGTAGCATCAGGATCGCGCGCGCAGCTCAAGAGGCGCTGTGTGACACCG 300  
 Qy 458 CCGACGACACCGTGTGTGAGGAGATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCG 517  
 Db 301 CCGACGACACCGTGTGTGAGGAGATGAACCTGCGCGCGCAAGTGAAGCCCAAGATGATCG 360  
 Qy 518 GCGGCATCGCGGGCTTCAAGGTGGCGCGCAGTACGACACAGATCCTGATCGAGATCTGCG 577  
 Db 361 GCGGCATCGCGGGCTTCAAGGTGGCGCGCAGTACGACACAGATCCTCGTGAGATCTGCG 420  
 Qy 578 GCAAGAGGCGCATCGGCACCGTGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637  
 Db 421 GCAAGAGGCGCATCGGCACCGTGTGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCA 480  
 Qy 638 ACATGTGACCCAGCTGGGTGCACTGTGAACTTTCCCATCAGCCCCATCAGAGACCGTGC 697  
 Db 481 ACCTGTGACCCAGTCTGGCTGCACCTGAACTTTCCCATCAGCCCCATCAGAGACCGTGC 540  
 Qy 698 CCGTGAAGTGAAGCCCGCGCATGGAAGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGG 757  
 Db 541 CCGTGAAGTGAAGCCCGCGCATGGAAGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGG 600  
 Qy 758 AGAAGATCAAGGGCCCTGACCGCATCTGCGAGGAGATGGAAGAGGAGGGCAAGATCACCA 817  
 Db 601 AGAAGATCAAGGGCCCTGCTGAGATCTGACCGGAGATGGAAGAGGAGGGCAAGATCAGCA 660  
 Qy 818 AGATCGGCGCGCGAGAACCCCTCAACACCCCGGTGTTGCCATCAAGAGAGGACAGCA 877  
 |||||

```
|||||
661 AGATCGGCCCCGAGAACCCCTTAAACACCCCGCTGTTCCCATCAAGAGAAGGACAGCA 720
QY
878 CCAAGTGGCGCAAGCTGCTGGAGCTTCCGCGAGCTGAAACAAAGCGGACCCAGGACTTCTGGG 937
Db
721 CCAAGTGGCGCAAGCTGCTGGAGCTTCCGCGAGCTGAAACAAAGCGGACCCAGGACTTCTGGG 780
QY
938 AGGTGAGCTGGGCAATCCCAACCCCGCGGCTGAGAGAGAGAGAGAGGCTGACCGTGC 997
Db
781 AGGTGAGCTGGGCAATCCCAACCCCGCGGCTGAGAGAGAGAGAGAGGCTGACCGTGC 840
QY
998 TGGACGTGGGCGAGCGCTACTTTCAGGGCTGCCCTTGGACGAGGACTTCCCGCAAGTACACCG 1057
Db
841 TGGACGTGGGCGAGCGCTACTTTCAGGGCTGCCCTTGGACGAGGACTTCCCGCAAGTACACCG 900
QY
1058 CTTTCAACCATCCCAAGCATCAACAAAGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGC 1117
Db
901 CTTTCAACCATCCCAAGCATCAACAAAGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGC 960
QY
1118 TGGCCCGAGGGCTGGAGGGGAGCCCGAGATCGTGATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db
961 TGGCCCGAGGGCTGGAGGGGAGCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
QY
1178 AGCCCTTTCGCGCGCCGCAACCCCGAGATCGTGATCTTCCAGAGCAGCATGACCAAGATCCTGG 1237
Db
1021 AGCCCTTTCGCGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1080
QY
1238 GCGACCTGGAGATCGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1297
Db
1081 GCGACCTGGAGATCGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1140
QY
1298 GCTGGGGCTTACACACCCCGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1357
Db
1141 GCTGGGGCTTACACACCCCGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1198
QY
1358 GCTACAGCTGCACCCCGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1417
Db
1199 -----CGAGCTGCACCCCGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1254
QY
1418 GCTGGACCGTGAACGACATCCAGAGCTGTGGGCGAGCTGAACTGGGCGCAGCAGATCT 1477
Db
1255 GCTGGACCGTGAACGACATCCAGAGCTGTGGGCGAGCTGAACTGGGCGCAGCAGATCT 1314
QY
1478 ACCCGCGCATCAAGGTGCGGCGAGCTGTGCAAGCTGTGCGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1537
Db
1315 ACGCCGCGCATCAAGGTGAAGAGCTGTGCAAGCTGTGCGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1374
QY
1538 ACATCGTGCCTTACCGAGGAGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCCTGC 1597
Db
1375 AGGTGATCCCTTACCGAGGAGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCCTGA 1434
QY
1598 GCGAGCCCGTGCACGCGGTGTACTACGACCCGAGAGGACCTGGTGGCGGAGATCCAGA 1657
Db
1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCGAGAGGACCTGGTGGCGGAGATCCAGA 1494
QY
1658 AGCAGGGCCACGACGAGTGAACCTACAGATCTACAGGAGCCCTTCAAGAACTGAAGA 1717
Db
1495 AGCAGGGCCACGAGGCGAGTGAACCTACAGATCTACAGGAGCCCTTCAAGAACTGAAGA 1554
QY
1718 CCGGCAAGTACGCGCAAGATCGCGACCGCCACACCAACGACGCTGAAGCAGCTGACCGAGG 1777
Db
1555 CCGGCAAGTACGCGCGCATCGCGGCGCCACACCAACGACGCTGAAGCAGCTGACCGAGG 1614
QY
1778 CCGTGCAGAAAGATCGCCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGC 1837
Db
1615 CCGTGCAGAAAGTGAACCGAGAGCATCTGTGATCTGGGGCAAGATCCCAAGTTCGAAGC 1674
QY
1838 TGGCCCATCCAGAGGAGACCTGGAGACCTGGTGGACCGACCTACTGCGAGGCCACCTGGA 1897
Db
1675 TGGCCCATCCAGAGGAGACCTGGAGGCGCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1734
QY
1898 TCCCGGAGTGGGAGTTCTGTAACACACCCCGCTTGGTGAAGCTGTGGTACAGCTGGAGA 1957
|||||
```

```
Db
1735 TCCCCGAGTGGGAGTTCTGTGAACACACCCCGCTGTGTGAAGCTGTGTGTAACCGAGTGGAGA 1794
QY
1958 AGGAGCCCATCATCGGCGCGGAGACCTTCTAGTGAACGCGCGCCCAACCGGAGACCA 2017
Db
1795 AGGAGCCCATCGTGGCGCGGAGACCTTCTAGTGAACGCGCGCCCAACCGGAGACCA 1854
QY
2018 AGATCGGCAAGGCGCGCTACGTGACCGAACCGGGCGCGGAGAAAGATCGTGAGCTCGACCG 2077
Db
1855 AGCTGGGCAAGCGCGCTAGTGAACCGCGCGCGGCGGAGAGGTGCTGAGCATCGCGC 1914
QY
2078 AGACCAACCAACCAAGAGACCGAGCTGCGAGCCATCCAGCTGGCCCTGCGAGGACAGCGGCA 2137
Db
1915 ACACCAACCAACCAAGAGACCGAGCTGCGAGCCATCCACCTGGCCCTGCGAGGACAGCGGCC 1974
QY
2138 GCGAGGTGAACATCGTGAACCGGAGCCAGTACCGCTTGGGCATCCAGGCGCCAGCCCG 2197
Db
1975 TGGAGGTGAACATCGTGAACCGGAGCCAGTACCGCTTGGGCATCCAGGCGCCAGCCCG 2034
QY
2198 ACAAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAAGG 2257
Db
2035 ACNAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAGAGGAGAGG 2094
QY
2258 TGTACCTGAGCTGGTGGCGGCCCAAGAGGCGATCGCGGCGAACAGAGCAGATCGACAAGC 2317
Db
2095 TGTACCTGAGCTGGTGGCGGCCCAAGAGGCGATCGCGGCGAACAGAGCAGTGGACAAGC 2154
QY
2318 TGGTGAACGAGGCGATCCGCGAGTGTCTTCTGGAACGCGGCGCTAGGATCGATTAAGC 2377
Db
2155 TGGTGAACGCGCGCATCCGCGAGTGTCTTCTGGAACGCGGCGCTAGGATCGATTAAGC 2214
QY
2378 TCTACAGGTACATGGAACGACCTGTAGCTGGGCGAGCGCGGCGCTAGGATCGATTAAGC 2437
Db
2215 TCTACAGGTACATGGAACGACCTGTAGCTGGGCGAGCGCGGCGCTAGGATCGATTAAGC 2274
QY
2438 TCCCCGGGCTAGCACCGGTGAATTC 2463
Db
2275 TCCCCGGGCTAGCACCGGTGAATTC 2300

RESULT 14
AX427936
LOCUS
DEFINITION Sequence 174 from Patent WO0232943.
ACCESSION AX427936
VERSION AX427936.1 GI:21538023
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE
1 Huang, Y. and Nabel, G.J.
AUTHORS Modifications of hiv env, gag, and pol enhance immunogenicity for
TITLE Genetic immunization
JOURNAL Patent: WO 0232943-A 174 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source Location/Qualifiers
1..9788
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 82.0%; Score 2019.2; DB 6; Length 9788;
Best Local Similarity 91.6%; Pred. No. 4.8e-202;
Matches 2190; Conservative 0; Mismatches 178; Indels 24; Gaps 4;

QY 14 TGGCCCGAGGCGCATGAGCAGGCGCCAGCGCCCAACATCTGTGATGCGAGCGCAGCAACTCA 73
Db 2961 TGGCCCGAGGCGCATGAGCAGGCGTGAACACCAACCATCATGATGCGAGCGCAGCACTGCA 3020
QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCGAAGGAGGCGCACATCGCCCGCA 133
|||||
```





```
QY 2276 CCGCCCAAGGGCATCGCGGCAAGCAGCAGATCGACAGCTGGTGCAGCAAGGGCATCC 2335
Db |||||
5235 CCGCCCAAGGGCATCGCGGCAAGCAGCAGTGGACGGCTGGTGAGCGCGGCATCC 5294
QY 2336 GCAAGTGCTGTTCTTGAGCGGCATCGATGGCGGCATCGTGATCTACCACTGA 2387
Db |||||
5295 GCAAGTGCTGTTCTTGAGCGGCATCGACAAAGGCCCGAGGAGGAGCAGAGAA 5346

RESULT 15
AX427931
LOCUS AX427931 9169 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 169 from Patent WO0232943.
ACCESSION AX427931
VERSION AX427931.1 GI:21538018
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic HIV genes.
other sequences; artificial sequences.
REFERENCE
1 Huang Y. and Nabel G.J.
AUTHORS Modifications of hiv env, gag, and pol enhance immunogenicity for
TITLE Genetic immunization
JOURNAL Patent: WO 0232943-A 169 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1. 9169
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 81.4%; Score 2005.2; DB 6; Length 9169;
Best Local Similarity 91.4%; Pred. No. 1.4e-200;
Matches 2187; Conservative 0; Mismatches 183; Indels 22; Gaps 5;

QY 14 TGCCCGAGGCGATGAGCGGCCACCAGCGGCCAACATCTCTGATGCGAGCGCAGCACTTCA 73
Db |||||
2967 TGCCCGAGGCGATGAGCGGCCACCAGCGGCCAACATCTCTGATGCGAGCGCAGCACTTCA 3026

QY 74 AGGCCCCCAAGCGCATCATCAAGTGTCTCAATGCGCGCAAGGAGGCGCCACATCGCCGCA 133
Db |||||
3027 AGGCCCCAAGCGCAT---CAAGTGTCTCAACTGCGCGAAGGAGGCGCCACTGCGCCGCA 3083

QY 134 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAACTGGCGCAAGGAGGCGCCACCATGA 193
Db |||||
3084 ACTGCGCGCGCCTGCGCAAGAGGGCTGTGGAACTGGCGCAAGGAGGCGCCACCATGA 3143

QY 194 AGGACTGCACCGAGCGCGCCAACTTCTTCGCGAGGACCTGGCTTCCCCCAGGGCA 253
Db |||||
3144 AGGACTGCACCGAGCGCGCCAACTTCTTCGCGAGGAGATCTGGCTTCCCCCAGGGGA 3202

QY 254 AGGCGCGCGAGTTTCCCCAGCGAGCAGAACCGCGCCCAAGCCGCCACCGCGCGAGCTGC 313
Db |||||
3203 AGGCGCGAGGAAATTTCTTCAGAGCAGACAGAGCCAAAGCCCAAGAGAGAGCTTC 3262

QY 314 AGGTGCGCGG-----CGAACCCCGGAGCGAGGCGCGCGCGAGCGCGCA--- 364
Db |||||
3263 AGGTTTGGGGAAGAGAACAACTCCCTCTCAGAAAGCAGGAGCGCGATAGCAAGGAACTG 3322

QY 365 ---CCCTGAACTTCCCCAGATCACTGTGGCAGCGCCCTTGGTGAGCATCAAGGTGG 421
Db |||||
3323 TATCTTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACATTAAGATAG 3382

QY 422 GCGGCGAGATCAAGAGGCGCCTGTGAGACACCGGCGCGCAGACCGGTGCTGAGGAGA 481
Db |||||
3383 GGGGCGAGCTGAAGGAGGCGCTTCTAGACACCGGCGCGCAGACACCGGTGCTGAGGAGA 3442

QY 482 TGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGGGCTTTCATCAGG 541
Db |||||
3443 TGAACCTGCCCGCGCTGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAGG 3502
```

```
QY 542 TGGCCGAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGC 601
Db |||||
3503 TGGCCGAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGC 3562

QY 602 TGAATCGGCCCCACCCCGGTGAACATCATCGGCGCGCAACATGCTGAGCCAGCTGGGCTGCA 661
Db |||||
3563 TGGTGGGCCCCACCCCGGTGAACATCATCGGCGCGCAACATGCTGAGCCAGATCGGCTGCA 3622

QY 662 CCTGAATCTTCCCATCAGCCCATCGAGACCGTGCCTGAACTGAAAGCCCGGATGG 721
Db |||||
3623 CCTGAATCTTCCCATCAGCCCATCGAGACCGTGCCTGAACTGAAAGCCCGGATGG 3682

QY 722 ACGGCCCCAAGGTGAAGCAGTGCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781
Db |||||
3683 ACGGCCCCAAGGTGAAGCAGTGCCTGACCGAGGAGAAGATCAAGGCCCTGAGTGA 3742

QY 782 TCTGCGAGGAGATGGAGAAAGAGGGGCAAGATCAACAAAGATCGGCCCGGAGAACCCCTACA 841
Db |||||
3743 TCTGCAACCGAGATGGAGAAAGAGGGGCAAGATCAACAAAGATCGGCCCGGAGAACCCCTACA 3802

QY 842 ACACCCCGCTGTTCGCCCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTGACT 901
Db |||||
3803 ACACCCCGCTGTTCGCCCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTGACT 3862

QY 902 TCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGCATCCCCACC 961
Db |||||
3863 TCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGCATCCCCACC 3922

QY 962 CGCGCGGCTGAAGAAAGAGAGCGTGCACGTGCTGACGTGGAGCTGGGCGAGCCCTACTTCA 1021
Db |||||
3923 CGCGCGGCTGAAGAAAGAGAGCGTGCACGTGCTGGAAGCTGGGCGAGCCCTACTTCA 3982

QY 1022 GGTGCGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGCAGCATCAACA 1081
Db |||||
3983 GGTGCGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGCAGCATCAACA 4042

QY 1082 ACAGAGACCCCGGCTACCGTACCAAGTACAAAGTGTGCTGCCCAAGGCTGGAAGGGAGCC 1141
Db |||||
4043 ACAGAGACCCCGGCTACCGTACCAAGTACAAAGTGTGCTGCCCAAGGCTGGAAGGGAGCC 4102

QY 1142 CMAGCATCTTCAGAGCAGATGACCAAGATCTTGAGCGCTTCCGCGCCCGCAACCCCG 1201
Db |||||
4103 CGGCCATCTTCCAGTCAGCATGACCAAGATCTTGAGCGCTTCCGCAAGAGAGAGAGCCCG 4162

QY 1202 AGATCGTGTATCTACCA---GGCCCCCTGTACGTGGCGCAGCAGCTTGGAGATCGGCC 1255
Db |||||
4163 ACATCGTGTATCTACCAAGTACATGGAACACTGTACGTGGCGCAGCAGCTTGGAGATCGGCC 4222

QY 1256 AGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTGTGCTGGGCTTCAACACCC 1315
Db |||||
4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTGCTGGGCTTCAACACCC 4282

QY 1316 CCGAAGAGAGCAACAGAGGAGCCCTTCTCTGTGATGGGCTAGGCTGACCCCG 1375
Db |||||
4283 CCGAAGAGAGCAACAGAGGAGCCCTTCTCTGTGATGGGCTAGGCTGACCCCG 4342

QY 1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAAAGAGAGCTGACCGTGAACGACA 1435
Db |||||
4343 ACAAGTGGACCGTGCAGCCCATCGTGTGCCGAGAAAGAGAGCTGGAACGCTGAACGACA 4402

QY 1436 TCAGAGAGCTGGTGGCAAGCTGAACTGGGCCAGCGAGATCTATCCCGCGCATCAAGGTGC 1495
Db |||||
4403 TCAGAGAGCTGGTGGCAAGCTGAACTGGGCCAGCGAGATCTATCCCGCGCATCAAGGTGC 4462

QY 1496 GCAGCTGTGAAGCTGTGCGCGGCGCCAGGCCCTTGACCGCATGCTGCCCTGACCC 1555
Db |||||
4463 GCAGCTGTGAAGCTGTGCGCGGCGCCAGGCCCTTGACCGCATGCTGCCCTGACCC 4522

QY 1556 AGGAGCCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCAGCGCCGTGACGGCG 1615
Db |||||
4523 AGGAGCCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGAGAGGAGCCGTGACGGCG 4582

QY 1616 TGTACTACGACCCAGCAAGGAGCTTGGTGGCGAGATCCAGAAAGCAGGGGCCACGACAGT 1675
```

[illegible]

Search completed: June 1, 2005, 20:20:37  
Job time : 10550.7 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 09:47:50 ; Search time 389.948 Seconds  
(without alignments)  
10360.282 Million cell updates/sec

Title: US-09-610-313B-30  
Perfect score: 2469  
Sequence: 1 gtcagccaccatgccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 résidues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgm2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgm2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgm2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgm2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgm2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgm2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2052	83.1	2312	4	US-09-475-515-84
2	2025.2	82.0	2306	4	US-09-475-515-82
3	1998.4	80.9	2300	4	US-09-475-515-83
4	1969.4	79.8	4319	4	US-09-475-515-6
5	1858	75.3	2305	4	US-09-475-515-80
6	1831.2	74.2	2299	4	US-09-475-515-81
7	1678.6	58.0	4307	4	US-09-552-950-2
8	1659.4	67.2	4307	4	US-09-936-572-2
9	1651.4	66.9	4327	4	US-09-936-572-14
10	1651.4	66.9	4353	4	US-09-936-572-13
11	1651.4	66.9	4642	4	US-09-936-572-12
12	1651.4	66.9	9772	4	US-09-552-950-5
13	1582.8	64.5	8366	4	US-09-872-733A-6
14	1557.4	63.1	4338	4	US-09-872-733A-1
15	1336.4	54.1	2577	4	US-09-952-060-1
16	1336.4	54.1	2650	4	US-09-952-060-5
17	1327.4	53.8	4053	4	US-09-952-060-34
18	1325.2	53.7	2577	4	US-09-952-060-3
19	1325.2	53.7	2650	4	US-09-952-060-7
20	1325.2	53.7	38519	4	US-09-952-060-28
21	1227.2	49.7	9010	4	US-09-184-418C-8
22	1199.4	48.6	9913	4	US-09-827-688-11
23	1196.2	48.4	8972	4	US-09-184-418C-9
24	1190.6	48.2	2467	4	US-09-872-733A-3
25	1189	48.2	8959	4	US-09-184-418C-11
26	1165.6	47.2	8992	4	US-09-184-418C-4
27	1132	45.8	2601	3	US-09-117-217-7

Sequence 9, Appli  
Sequence 11, Appl  
Sequence 13, Appl  
Sequence 7, Appli  
Sequence 9, Appli  
Sequence 11, Appl  
Sequence 13, Appl  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 7, Appli  
Sequence 9, Appli  
Sequence 1, Appli  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 15, Appl  
Sequence 13, Appl

## ALIGNMENTS

## RESULT 1

US-09-475-515-84  
; Sequence 84, Application US/09475515A  
; Patent NO. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 84  
; LENGTH: 2312  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS (-).protmod.RTopt (+)  
US-09-475-515-84

Query Match 83.1%; Score 2052; DB 4; Length 2312;  
Best Local Similarity 93.6%; Pred. No. 0;  
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
  
Qy 170 GCGCAAGGAGGGCCACGAGTGAAGGACTGCAGCGCGCGAGCCAGCTTCTTCGCG 229  
Db 1 CGGCGCGGAGGAGGACCAATGAAGATTGCACTGAGAGAGGCTAATTTCTTCGCG 60  
  
Qy 230 AGGACCTGGCTTTCCCGCGAGGCGCGAGTTCCCGAGCGAGCAGAACCGCGCA 289  
Db 61 AGGACCTGGCTTCTTCGAGGCAAGCGCCGCGAGTTTCAGCAGCGAGCAGACCGCGCA 120  
  
Qy 290 ACAGCCCCACAGCGCGAGCTGCGCGCG-----ACAACCCCGAGCGAGG 343  
Db 121 ACAGCCCCACCGCGCGAGCTGCGAGTGTGGGCGCGGAGAACAGAGCTGAGCGAGG 180  
  
Qy 344 CCGCGCGCGAGCGCGAGGCGACCGCTG-----AACTTCCCGCAGATCACCTGTGGCAGC 397  
Db 181 CCGCGCGCGAGCGCGAGGCGACCGCTGAGCTTCACTTCCCGCAGATCACCTGTGGCAGC 240



```

: APPLICANT: BARNETT, Susan
: APPLICANT: ZUR MEGEDE, Jan
: APPLICANT: SRIVASTAVA, Indresh
: APPLICANT: LIAN, Ying
: APPLICANT: HARTOG, Karin
: APPLICANT: LIU, Hong
: APPLICANT: GREER, Catherine
: APPLICANT: SELBY, Mark
: APPLICANT: WALKER, Christopher
: TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
: TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
: FILE REFERENCE: 1621.002
: CURRENT APPLICATION NUMBER: US/09/475,515A
: CURRENT FILING DATE: 1999-12-30
: NUMBER OF SEQ ID NOS: 90
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 82
: LENGTH: 2306
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence:
: OTHER INFORMATION: FS(-), protmod.RTopt.YM
US-09-475-515-82

Query Match      82.0%; Score 2025.2; DB 4; Length 2306;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

Qy      170  GCGCGCAGGAGGGCCACAGATGAAGACTGCACCGAGCGCCAGGCCAACTCTTCGCGG 229
Db      1   GCGCGCGGAGGAGCACCAATGAAGATTGCACTGAGAGACAGGCTAAATTTCTTCGCGG 60

Qy      230  AGGACCTGGCCCTCCCCAGGGCAAGGCCGCGAGTTCCCCAGCGAGCAGAGAACCGCGCCA 289
Db      61  AGGACCTGGCCCTTCTGAGGGCAAGGCCCGCGAGTTGAGCAGCGAGCAGACCCGCGCCA 120

Qy      290  ACAGCCCCACACGCGCGAGCTGCAGGTGCGCGCG-----ACAAACCCCGCGAGCGAGG 343
Db      121  ACAGCCCCACCCCGCGCGAGCTGCAGGTGCGCGCGCGAGAACAAACAGCCTGAGCGAGG 180

Qy      344  CCGCGCGCGAGCGCCAGGGCAACCTG-----AAGTCCCCAGATCAACCTGTGCGAGC 397
Db      181  CCGCGCGCGACCGCGCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCCCTGTGGCAGC 240

Qy      398  GCCCCCTGTGTAGACATCAAGGTGGCGCGCAGATCAAGGAGGCCCTGCTGGACACCGCGG 457
Db      241  GCCCCCTGTGTGACCATCAGGATCGCGCGCGCAGCTCAAGGAGGCGCTGCTCGACACCGCGG 300

Qy      458  CCGACGACACCGTGTGTGAGGAGATGAGCCTTCCCGCGCAAGTGGAAAGCCCAAGATGATCG 517
Db      301  CCGACGACACCGTGTGTGAGGAGATGAACCTGCCCGCAAGTGGAAAGCCCAAGATGATCG 360

Qy      518  GCGGCAATCGGCGGCTTCATCAAGGTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCG 577
Db      361  GCGGCAATCGGCGGCTTCATCAAGGTGCGCGAGTACGACAGATCCCGTGGAGATCTGCG 420

Qy      578  GCAGAGGGCCATCGGCACCGTGTGATCGCGGCCCAACCCCGTGAAACATCATCGGCGCGCA 637
Db      421  GCACAAGGGCCATCGGCACCGTGTGTGTGGGCGCCCAACCCCGTGAACATCATCGGCGCGCA 480

Qy      638  ACATGCTGACCCAGCTGGGGCTGCACCCCTGAACCTTCCCCCATCAGGCCCATCGAGACCGTGC 697
Db      481  ACCTGCTGACCCAGATCGGCTGCACTTGAACCTTCCCCCATCAGGCCCATCGAGACCGTGC 540

Qy      698  CCGTGAAGCTGAAGCCCGGCATGAGACGCGCCCCAAGGTGAAGCAGTGGCGCCCTTGACCGAGG 757
Db      541  CCGTGAAGCTGAAGCCCGGGATGAGCGCGCCCCAAGGTCAAGCAGTGGCGCCCTTGACCGAGG 600

Qy      758  AGAAGATCAAGGCCCTGACCCCATCTTGCAGAGGAGATGGAGAAAGGAGGGCAAGATCAACCA 817
Db      601  AGAAGATCAAGGCCCTTGTGTGAGATCTGTCCACCGAGATGGAGAAAGGAGGGCAAGATCAACA 660

```



```

Db 1735 CCTGGATCCCGAGTGGAGTCTGTGAACACCCCCCTGTGTGAAGCTGTGTGATCACCG 1794
QY 1958 TGAGAGAGAGCCCATCATCGGCGCGAGACCTTCTAGCTGAGACGGCGCGCCAAACCGCG 2017
Db 1795 TGAGAGAGAGAGCCCATCGTGGCGCGAGACCTTCTAGCTGAGACGGCGCGCCAAACCGCG 1854
QY 2018 AGACCAAGATCGCAAGCGCGCTAGCTGACCGACCGGGCGCGCAGAGATCGTGAGCC 2077
Db 1855 AGACCAAGCTTGGCAAGCGCGCTAGCTGACCGACCGGGCGCGCAGAGAGTGGTGAGCA 1914
QY 2078 TGACCGAGACCAACCAACAGACCGAGCTGACCGACCGAGCTGAGCTGGCCCTCGCAGGACA 2137
Db 1915 TGCCCGACACCAACCAAGACCGAGCTGACCGACCGGGCGCGCAGAGAGTGGTGAGCA 1974
QY 2138 GCGGACGAGAGTGAACATCGTGACCGACAGCCAGTACCGCTTGGGGATCATCCAGGCC 2197
Db 1975 GCGGCTTGGAGTGAACATCGTGACCGACAGCCAGTACCGCTTGGGGATCATCCAGGCC 2034
QY 2198 AGCCCGACAGAGCGAGAGCGAGCTGTGTGAACCGACAGATCATCGAGAGCTGATCAAGAAG 2257
Db 2035 AGCCCGACAGAGCGAGAGCGAGCTGTGTGAACCGACAGATCATCGAGAGCTGATCAAGAAG 2094
QY 2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCACCAAGGCGCATCGGCGGCAACGAGCAGATCG 2317
Db 2095 AGAAGGTGTACCTGGCTGGGTGCCCGCCACCAAGGCGCATCGGCGGCAACGAGCAGTGG 2154
QY 2318 ACAAGCTGTGAGCAAGGGGATCCGCAAGGTGTGTCTTGTGAGCGGATCGATGGCGGCA 2377
Db 2155 ACAAGCTGTGAGCGCGGATCCGCAAGGTGTGTCTTGTGAAACGGCATCGATGGCGGCA 2214
QY 2378 TCGTGATCTACAGTACATGAGAGCTGTGACCTGTGGGCGAGCGGCGCCCTAGGATCGATT 2437
Db 2215 TCGTGATCTACAGTACATGAGAGCTGTGACCTGTGGGCGAGCGGCGCCCTAGGATCGATT 2274
QY 2438 AAAAGCTTCCCGGGCTAGCACCGGTGAATTG 2469
Db 2275 AAAAGCTTCCCGGGCTAGCACCGGTGAATTG 2306

```

```

RESULT 3
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475, 515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: PS(-).protmod.RTopt.YMMW
US-09-475-515-83

```

```

Query Match 80.9%; Score 1998.4; DB 4; Length 2300;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

```

```

QY 170 GCGCAAGAGGCGCCACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCG 229
Db 1 GCGCGCGAGAGGACACCAATGAAGATTGACTGAGAGACAGGCTAATTTCTTCGCG 60
QY 230 AGGACCTGCGCTTCCCGCAGGGCAAGGCGCGAGTTCCCGCAGCAGAGAAACCGCGCCA 289
Db 61 AGGACCTGCGCTTCCCGCAGGGCAAGGCGCGAGTTCCAGCAGCAGCAGACCCGCGCCA 120
QY 290 AAGCCCCACAGCCCGAGCTGCAAGTGGCGGGG-----ACAACCCCGCAGCGAG 343
Db 121 AAGCCCCACAGCCCGCGAGCTGCAAGTGGCGGGGCGAGAAACAACAGCTGAGCGAG 180
QY 344 CGGCGCGCAGCGCAGGCGCACCTG-----AACTTCCCGCAGATCACCTCTGGCAGC 397
Db 181 CGGCGCGCAGCGCAGGCGCACCTGAGCTTCACTTCCCGCAGATCACCTCTGGCAGC 240
QY 398 GCGCGCTGTGAGCATCAAGTGGCGCGCAGATCAAGGAGGCCCTGTGTGAACACCGCG 457
Db 241 GCGCGCTGTGAGCATCAAGTGGCGCGCAGCTCAAGGAGGCCCTGTGTGAACACCGCG 300
QY 458 CGGACGACACCGCTGTGAGGAGATGAGCTGCGCGGCAAGTGGAGGCCCAAGATGATCG 517
Db 301 CGGACGACACCGCTGTGAGGAGATGAGCTGCGCGGCAAGTGGAGGCCCAAGATGATCG 360
QY 518 GCGCATCGCGGCTTTCATCAAGTGGCGCAGTACGACAGATCTCTGATCGAGATCTGCG 577
Db 361 GCGGATCGGGGCTTTCATCAAGTGGCGGAGTACGACAGATCTCCCGTGGAGATCTGCG 420
QY 578 GCAAGAGGCCATCGGACACCGTGTGATCGGCGCCACCCCGTGAACATCATCGGCGCA 637
Db 421 GCCACAAGGCCATCGGACACCGTGTGATCGGCGCCACCCCGTGAACATCATCGGCGCA 480
QY 638 ACATGCTGACCGAGCTGGGCTGCACTTCCCGATCAGCCCATCGAGACCGTGC 697
Db 481 ACCTGCTGACCGAGATCGGCTGCACCTTGAACCTTCCCGATCAGCCCATCGAGACCGTGC 540
QY 698 CGTGAAGCTGAAGCCCGCATGCGCGCCCAAGGTGAAGCAGTGGCGCCCTGACCGAG 757
Db 541 CGTGAAGCTGAAGCCCGGATGAGCGGCCCAAGGTGAAGCAGTGGCGCCCTGACCGAG 600
QY 758 AGAAGATCAAGGCCCTGACCGCCATCTGGAGAGATGGAGAAAGGAGGCCAAGATCACCA 817
Db 601 AGAAGATCAAGGCCCTGTTGGAGATCTGCACCGAGATGGAGAAAGGAGGCCAAGATCAGCA 660
QY 818 AGATCGGCGCGAGAACCCCTCAACAACCCCGTGTTCGCCATCAAGAGAGAGCAGCA 877
Db 661 AGATCGGCGCGAGAACCCCTCAACAACCCCGTGTTCGCCATCAAGAGAGAGCAGCA 720
QY 878 CCAAGTGGCGCAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db 721 CCAAGTGGCGCAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780
QY 938 AGTGCGAGCTGGGCTATCCCGACCCCGCGCTGAAGAGAGAGAGCGTGAACCGTGC 997
Db 781 AGTGCGAGCTGGGCTATCCCGACCCCGCGCTGAAGAGAGAGAGCGTGAACCGTGC 840
QY 998 TGGACGTGGCGAGCGCTACTTTCAGCGTGGCGCGCCCTGAGCAGGAGCTTCCGCAAGTACACCG 1057
Db 841 TGGACGTGGCGAGCGCTACTTTCAGCGTGGCGCGCCCTGAGCAGGAGCTTCCGCAAGTACACCG 900
QY 1058 CTTTACCATCCCCAGCATCAACAAGCAGACCCCGCGCATCCCGTACCAGTACAAAGTGC 1117
Db 901 CTTTACCATCCCCAGCATCAACAAGCAGACCCCGCGCATCCCGTACCAGTACAAAGTGC 960
QY 1118 TGCCCCAGGGCTGGAAGGGCAGCCCGAGCATTTCCAGAGCAGCATGACCAAGATCTGG 1177
Db 961 TGCCCCAGGGCTGGAAGGGCAGCCCGCGCATTTCCAGAGCAGCATGACCAAGATCTGG 1020
QY 1178 AGCCCTTCCGCGCGCCCAAGATCGTGTATCTTACAGTACATGAGCAGCTGATCG 1237
Db 1021 AGCCCTTCCGCGCGCCCAAGATCGTGTATCTTACAGTACATGAGCAGCTGATCG 1074

```



Qy	419	TGGCGCGCCAGATCAAGAGGSCCTGCTGGACACCGCGCGCGAGCACCGTGTCTGGAGG	478
Db	1519	TCGGCGCGCAGCTCAAGAGGCGCTGTGTGACACCGCGCGCGAGCACCGTGTCTGGAGG	1578
Qy	479	AGATGAGCTCTCCGGCNAAGTGGAGGCCCAAGATGATCGGGCGCATCGGCGGCTTCATCA	538
Db	1579	AGATGAACCTTCCCGGCAAGTGGAAAGGCCAAGATGATCGCGCGGATCGGGGGCTTCATCA	1638
Qy	539	AGGTGCGCCAGTACGACAGATTCCTGATCGAGATCTCGGCAAGAAAGGCCATCGGCACCG	598
Db	1639	AGGTGCGCGAGTACGACAGATCCCCTGAGATCTCGGCGCACAAGGCCATCGGCACCG	1698
Qy	599	TGCTGATCGGCCCAACCCCGTGAAACATCATCTCGGCGCGAAACATGTCTGACCCAGCTGGGCT	658
Db	1699	TGCTGTGTGGGCCCAACCCCGTGAAACATCATCTCGGCGCGAAACCTGCTGACCCAGATCGGCT	1758
Qy	659	GCACCTCTGAATCTCCCATCAGCCCCATCAGAGCCGTGCCCCGTGAAGCTCAAGCCCGGCA	718
Db	1759	GCACCTCTGAATCTCCCATCAGCCCCATCAGAGCGTGCCTGTGAAGCTGAAAGCCGGGA	1818
Qy	719	TGAGCGGCCCAAGGTGAAGCAGTGGCCCTGTACCGAGGAGAAAGATCAAGGCCCTGACCG	778
Db	1819	TGAGCGGCCCAAGGTCAAGCAGTGGCCCTGTACCGAGGAGAAAGATCAAGGCCCTGTGTG	1878
Qy	779	CCATCTGGAGGAGATGGAGAGGGCGAAGATACCAAGATCGGCCCCCGAGAAACCCCT	838
Db	1879	AGATCTGCACCGAGATGGAGAGGGCGAAGATCAGAAAGATCGGCCCCCGAGAAACCCCT	1938
Qy	839	ACAACACCCCGTGTTCGCCATCAAGAAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGG	898
Db	1939	ACAACACCCCGTGTTCGCCATCAAGAAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGG	1998
Qy	899	ACTTCCGCGAGCTCAACAAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC	958
Db	1999	ACTTCCGCGAGCTGAAACAAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC	2058
Qy	959	ACCCCGCGGCTGAAGAAGAAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACT	1018
Db	2059	ACCCCGCGGCTGAAGAAGAAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACT	2118
Qy	1019	TCAGCTGCCCCCTCGACGAGGACTTCCGCAAGTACCGGCTTCCACCATCCCAGAGATCA	1078
Db	2119	TCAGCTGCCCCCTCGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCAGAGATCA	2178
Qy	1079	ACAACGAGACCCCGGCATCCGTTACAGTACAAAGTGTGCCCCAGGGCTGAAAGGCA	1138
Db	2179	ACAACGAGACCCCGGCATCCGTTACAGTACAAAGTGTGCCCCAGGGCTGAAAGGCA	2238
Qy	1139	GCCCCAGCATTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGAACC	1198
Db	2239	GCCCCAGCATTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGAACC	2298
Qy	1199	CCGAGATCGTGATCTACCAAGTACATGACACCTGTGTACGTGGGCGAGGCACTCGAGATCG	1258
Db	2299	CCGACATCGTGATCTACCAAGTACATGACACCTGTGTACGTGGGCGAGGCACTCGAGATCG	2358
Qy	1259	GCCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTCAACCA	1318
Db	2359	GCCAGCACCGCACCAAGATCGAGAGCTGCGCGCAGCACCTGCTGCGTGGGGCTTCAACCA	2418
Qy	1319	CCCCCGACAGAAGCACCAAGAGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCACC	1378
Db	2419	CCCCCGACAGAAGCACCAAGAGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCACC	2478
Qy	1379	CCGACAAAGTGAACCGTGCAGGCCATCGAGCTGCCCGGAGAGGAGCTTGACCGTGAACG	1438
Db	2479	CCGACAAAGTGAACCGTGCAGGCCATCATGCTGCCCGGAGAGGAGCTTGACCGTGAACG	2538
Qy	1439	ACATCCAGAAGCTGGTGGCGAAGCTGAACTGGGCGACGCAAGATCTACCCCGGCATCAAGG	1498
Db	2539	ACATCCAGAAGCTGGTGGCGAAGCTGAACTGGGCGACGCAAGATCTACCCCGGCATCAAGG	2598

QY	1499	TGCGCAGCTGTGCAAGCTGCTGGCGGCGCCCAAGGCCCTGACCGACATCGTGCCTCTGA	1550	TGCGCAGCTGTGCAAGCTGCTGGCGGCGCCCAAGGCCCTGACCGACATCGTGCCTCTGA
DB	2599	TGAAGCAGCTGTGCAAGCTGCTGGCGGCGCCCAAGGCCCTGACCGAGGTGATCCCTCTGA	2650	TGAAGCAGCTGTGCAAGCTGCTGGCGGCGCCCAAGGCCCTGACCGAGGTGATCCCTCTGA
QY	1559	CCGAGGAGCGCAGCTGAGCTGGCCGAGAACCGCAGATGCTCTGGCGGAGCCCGGTGCACG	1610	CCGAGGAGCGCAGCTGAGCTGGCCGAGAACCGCAGATGCTCTGGCGGAGCCCGGTGCACG
DB	2659	CCGAGGAGCGCAGCTGAGCTGGCCGAGAACCGCAGATGCTCTGGCGGAGCCCGGTGCACG	2710	CCGAGGAGCGCAGCTGAGCTGGCCGAGAACCGCAGATGCTCTGGCGGAGCCCGGTGCACG
QY	1619	GCGTGTACTACGACCCGACAAAGACCTGTGTGGCGAGATCCAGAGCAGGGGCCACGACC	1670	GCGTGTACTACGACCCGACAAAGACCTGTGTGGCGAGATCCAGAGCAGGGGCCACGACC
DB	2719	AGGTGTACTACGACCCGACAAAGACCTGTGTGGCGAGATCCAGAGCAGGGGCCAGGGCC	2770	AGGTGTACTACGACCCGACAAAGACCTGTGTGGCGAGATCCAGAGCAGGGGCCAGGGCC
QY	1679	AGTGGACCTACAGATCTACAGAGGCCCTTCAAGAACCTGAAGCGGCAAGTACGCCCA	1730	AGTGGACCTACAGATCTACAGAGGCCCTTCAAGAACCTGAAGCGGCAAGTACGCCCA
DB	2779	AGTGGACCTACAGATCTACAGAGGCCCTTCAAGAACCTGAAGCGGCAAGTACGCCCC	2830	AGTGGACCTACAGATCTACAGAGGCCCTTCAAGAACCTGAAGCGGCAAGTACGCCCC
QY	1739	AGATGCGCACCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCG	1790	AGATGCGCACCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCG
DB	2839	GCATGCGCGCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGTGA	2890	GCATGCGCGCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGTGA
QY	1799	CCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAGG	1850	CCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAGG
DB	2899	GCACCGAGAGCATCTGTGATCTGGGGCAAGATCCCAAGTTCAAGCTGCCCATCCAGAAGG	2950	GCACCGAGAGCATCTGTGATCTGGGGCAAGATCCCAAGTTCAAGCTGCCCATCCAGAAGG
QY	1859	AGACCTGGGAGACCTGTGTGACCGACCTACTTGGCAGGCGCACCTGGATCCCCGAGTGGGAGT	1910	AGACCTGGGAGACCTGTGTGACCGACCTACTTGGCAGGCGCACCTGGATCCCCGAGTGGGAGT
DB	2959	AGACCTGGGAGGCTGTGTGATGTGGAGTACTTGGCAGGCCACCTGGATCCCCGAGTGGGAGT	3010	AGACCTGGGAGGCTGTGTGATGTGGAGTACTTGGCAGGCCACCTGGATCCCCGAGTGGGAGT
QY	1919	TCGTGAACAACCCCCCTGGTGAAGCTGTGTGTAACAGCTGTGAAGAGGAGCCCATCATCG	1970	TCGTGAACAACCCCCCTGGTGAAGCTGTGTGTAACAGCTGTGAAGAGGAGCCCATCATCG
DB	3019	TCGTGAACAACCCCCCTGGTGAAGCTGTGTGTAACAGCTGTGAAGAGGAGCCCATCTGTG	3070	TCGTGAACAACCCCCCTGGTGAAGCTGTGTGTAACAGCTGTGAAGAGGAGCCCATCTGTG
QY	1979	GGCGCAGAGACCTTCTAGTGTGACGGCGCCCAACCGCAGAGACCAAGATCGGCAAGGCCG	2030	GGCGCAGAGACCTTCTAGTGTGACGGCGCCCAACCGCAGAGACCAAGATCGGCAAGGCCG
DB	3079	GGCGCAGAGACCTTCTAGTGTGACGGCGCCCAACCGCAGAGACCAAGCTGGGCAGGCGCG	3130	GGCGCAGAGACCTTCTAGTGTGACGGCGCCCAACCGCAGAGACCAAGCTGGGCAGGCGCG
QY	2039	GCTACGTGTACCGACCGGGGCGGCAGAAAGATCGTGAAGCTTGACCGAGACCAACCAACGAG	2090	GCTACGTGTACCGACCGGGGCGGCAGAAAGATCGTGAAGCTTGACCGAGACCAACCAACGAG
DB	3139	GCTACGTGTACCGACCGCGCGCCAGAAAGTGTGTAGCATCGCCGACACCAACCAACGAG	3190	GCTACGTGTACCGACCGCGCGCCAGAAAGTGTGTAGCATCGCCGACACCAACCAACGAG
QY	2099	AGACCGAGCTGCAGGCCATCCAGTGGCCCTGCAGGACAGGCGCAGAGGTGAACATCG	2150	AGACCGAGCTGCAGGCCATCCAGTGGCCCTGCAGGACAGGCGCAGAGGTGAACATCG
DB	3199	AGACCGAGCTGCAGGCCATCCAGTGGCCCTGCAGGACAGGCGCTGGAGGTGAACATCG	3250	AGACCGAGCTGCAGGCCATCCAGTGGCCCTGCAGGACAGGCGCTGGAGGTGAACATCG
QY	2159	TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGCCGACCAAGAGCGAGAGCG	2210	TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGCCGACCAAGAGCGAGAGCG
DB	3259	TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGCCGACCAAGAGCGAGAGCG	3310	TGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGCCGACCAAGAGCGAGAGCG
QY	2219	AGCTGTGTAAACAGATCATCGACGCTGATCAAGAGGAGAGGTGTACTTGAGCTGGG	2270	AGCTGTGTAAACAGATCATCGACGCTGATCAAGAGGAGAGGTGTACTTGAGCTGGG
DB	3319	AGCTGTGTGAGCCAGATCATCGACGCTGATCAAGAGGAGAGGTGTACTTGAGCTGGG	3370	AGCTGTGTGAGCCAGATCATCGACGCTGATCAAGAGGAGAGGTGTACTTGAGCTGGG
QY	2279	TGCCCGCCCAACAGGGCATCGGGGCAACGAGCAGATTCGAACAGCTGGTGCAGCAAGGCGCA	2330	TGCCCGCCCAACAGGGCATCGGGGCAACGAGCAGATTCGAACAGCTGGTGCAGCAAGGCGCA
DB	3379	TGCCCGCCCAACAGGGCATCGGGGCAACGAGCAGAGGTGTGAACAGCTGGTGCAGCGCGCA	3430	TGCCCGCCCAACAGGGCATCGGGGCAACGAGCAGAGGTGTGAACAGCTGGTGCAGCGCGCA
QY	2339	TCGCCAAGGTGCTGTTCTTCGACGGGCATCGATGGCGGCATCGTGATCTACAGTGA	2390	TCGCCAAGGTGCTGTTCTTCGACGGGCATCGATGGCGGCATCGTGATCTACAGTGA
DB	3439	TCGCCAAGGTGCTGTTCTTCGACGGGCATCGACAGGCTCGAAGGCCAGGAGGAGCAACAGAA	3490	TCGCCAAGGTGCTGTTCTTCGACGGGCATCGAAGGCTCGAAGGCCAGGAGGAGCAACAGAA

RESULT 5  
US-09-475-515-80  
; Sequence 80, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine

APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475.515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 80  
LENGTH: 2305  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: PS(+).prolnact.RTopt.YM  
US-09-475-515-80

Query Match 75.3%; Score 1858; DB 4; Length 2305;  
Best Local Similarity 89.0%; Pred. No. 5.3e-309;  
Matches 2058; Conservative 0; Mismatches 235; Indels 19; Gaps 4;

Qy	170	GGGCAAGGAGGGCCACAGATGAAGACTGSCACGAGGCGCAGGCCAACTTCTTCGCG	229
Db	1	CGGCCCCGAGAGGACCAAAATGAAGATTGCATGAGACAGCGCTAAATTTT-AGGG	59
Qy	230	AGGACCTGGGCTTCCCCAGGCGAAGGCCGCGAGTTCCCCAGGAGCAGAACCGGCCA	289
Db	60	AAGATCTGGCTTCTTACAGGGAAGGCCAGGGAATTTCTTCAGAGCAGACGAGGCCA	119
Qy	290	ACAGCCCCACAGCCGCGAGTGGAGTGGCGCGCG-----ACAACCCCGCAGCGAGG	343
Db	120	ACAGCCCCACAGAGAGAGCTTCAGGTTTGGGAGGAGAAACAACTCCCTCTCAGAAG	179
Qy	344	CCGCGCGCGAGCGCCAGGCA-----CCCTGAACTTCCCGCAGATCACCCTGGCGAGC	397
Db	180	CAGAGCCGATAGACAAGGAATGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAAC	239
Qy	398	GCCCTCTGGTGAAGATCAAGGTGGCGGCCAGATCAAGGAGGCGCTGCTGGACACCGCG	457
Db	240	GACCCCTCGTACAAATGAAGATCGGGGGCACTCAAGGAAGCGCTGCTGATACAGGAG	299
Qy	458	CCGACGACACGCTGCTGGAGAGATGAGCTGCCCCGCAAGTGAAGCCCAAGATGATCG	517
Db	300	CAGATGATACAGTATTAGAAGAAATGAATTTGCGAGGAAATGGAACCAAAAATGATAG	359
Qy	518	CGGCGATCGGCGCTTCATCAAGTGGCGCAGTACGACGAGATCCTGATCGAGATCTGCG	577
Db	360	GGGGGATCGGGGGCTTTCATCAAGGTGAGGAGTACGACGACGATACCTGTAGAAATCTGTG	419
Qy	578	GCAAGAGGGCCATCGGCACCGTGTGATCGGCGCCACCGCGTGAACATCATCGCGCGCA	637
Db	420	GACATAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAATAATTTGAGAGAA	479
Qy	638	ACATGTGACCCAGCTGGGTGCAACCTTGAACTTTCCCTCATGCCCCCATCGAGACCGTGC	697
Db	480	ATCTGTTGACCCAGATCGGCTGCACTTGAATTTCCCTCATGCCCCCATTTGAGACGGTGC	539
Qy	698	CCGTGAAGCTGAGCCCGGCATGAGCGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGG	757
Db	540	CCGTGAAGTGAAGCCGGGATGAGCGGCCCCCAAGGTCAAGCAATGTGCCATTTGACCGAGG	599
Qy	758	AGAAGATCAAGGGCCCTGACCGCCATCTGCGAGGAGATGGAGAGGGGCAAGATCACCA	817
Db	600	AGAAGATCAAGGCCCTGTTGGAGATCTGCACCGGATGGAGAGGAGGGCAAGATCAGCA	659
Qy	818	AGATCGGCCCGAGAACCCCTTACAAACCCCGGTGTTGCGCATCAAGAAAGAGCAGCA	877
Db	660	AGATCGGCCCGAGAACCCCTTACAAACCCCGGTGTTGCGCATCAAGAAAGAGCAGCA	719
Qy	878	CCAAGTGGCGAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCAGCCAGGACTTCTGGG	937
Db	720	CCAAGTGGCGAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCAGCCAGGACTTCTGGG	779

Qy	938	AGGTGACAGCTGGGCATCCCCACCCCGCGCTTGAAGAAAGAAAGAGCGTGAACCGTGC	997
Db	780	AGGTGACAGCTGGGCATCCCCACCCCGCGCTTGAAGAAAGAAAGAGCGTGAACCGTGC	839
Qy	998	TGGAGCTGGGCGAGCGCTACTTTCAGCGTGGCCCTGGAGGAGGACTTCCGAAAGTACACCG	1057
Db	840	TGGAGCTGGGCGAGCGCTACTTTCAGCGTGGCCCTGGAGGAGGACTTCCGAAAGTACACCG	899
Qy	1058	CTTTCACCATCCCCAGCATCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1117
Db	900	CTTTCACCATCCCCAGCATCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	959
Qy	1118	TGCCCCAGGCTGGAAGGGCAGGCCAGCATTTTCAGAGAGCAGCATGACCAAGATCTCTGG	1177
Db	960	TGCCCCAGGCTGGAAGGGCAGGCCAGCATTTTCAGAGAGCAGCATGACCAAGATCTCTGG	1019
Qy	1178	AGCCCTTCCGCGCGCCGCAACCCCGAGATCGTGATCTACAGTACATGAGAGACCTGTACG	1237
Db	1020	AGCCCTTCCGCAAGAGCAACCCCGAGATCGTGATCTTACCA-----GGCCCCCTGTACG	1073
Qy	1238	TGGGACGACCTTGGAGATCGGCGCAGCAGCGCGCAAGATCGAGGAGCTGCGCAAGCACC	1297
Db	1074	TGGGACGACCTTGGAGATCGGCGCAGCAGCGCGCAAGATCGAGGAGCTGCGCAAGCACC	1133
Qy	1298	TGCTGGCTGGGCTTCAACACCCCGCAAGAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAG	1357
Db	1134	TGCTGGCTGGGCTTCAACACCCCGCAAGAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAG	1193
Qy	1358	GGATGGCTAGGAGCTGCACCCCGCAGAGAGTGGAGCGTGGAGCGCATCGAGTGGCCGAGA	1417
Db	1194	GGATGGCTAGGAGCTGCACCCCGCAGAGAGTGGAGCGTGGAGCGCATCATGTGCGCCGAGA	1253
Qy	1418	AGGAGAGCTGGACCGTGAACGACATCCAGAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCC	1477
Db	1254	AGGAGAGCTGGACCGTGAACGACATCCAGAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCC	1313
Qy	1478	AGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTGCTGCGGGCGCCCAAGGCCC	1537
Db	1314	AGATCTACCCCGCATCAAGTGAAGCAGCTGTGCAAGCTGCTGCGGGCGCCCAAGGCCC	1373
Qy	1538	TGACCGACATCTGTCCTCCCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAGACCGCAGA	1597
Db	1374	TGACCGAGGTGATCCCCCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAGACCGCAGA	1433
Qy	1598	TCCTGCGCAGCGCGTGCACGCGGTGATCTACGACCCAGCAGGAGCTGTGTGCGCCGAGA	1657
Db	1434	TCCTGAAGGAGCCCGTGCAAGAGTGTACTACGACCCAGCAGGAGAGAGAGAGAGAGAGAG	1493
Qy	1658	TCAGAGAGCGGCGCACGACAGTGGACCTTACAGATCTTACAGGAGCGCTTCAAGAAC	1717
Db	1494	TCAGAGAGCGGCGCACGAGGCGAGTGGACCTTACAGATCTTACAGGAGCGCTTCAAGAAC	1553
Qy	1718	TGAAGACCGCAAGTAGCGCCCAAGTGGCAGCGCCCGCCACCAACAGAGCTGAGCAGCTGA	1777
Db	1554	TGAAGACCGCAAGTAGCGCCCGAGTGGCAGCGCCCGCCACCAACAGAGCTGAGCAGCTGA	1613
Qy	1778	CCAGGCGCGTGCAGAGAGATCGCATGGAGAGAGTCTGTGATCTGGGGCAAGACCCCAAGT	1837
Db	1614	CCAGGCGCGTGCAGAGAGTGGACCGAGAGAGTCTGTGATCTGGGGCAAGATCTTCCCAAGT	1673
Qy	1838	TCCGCTTCCCATCCAGAGAGAGAGCTTGGAGAGAGCTTGGTGGAGCCGACTTCTGGCAGGCCA	1897
Db	1674	TCAAGCTGCCCATCCAGAGAGAGAGCTTGGAGAGAGCTTGGTGGAGTGTGTGGCAGGCCA	1733
Qy	1898	CTTGGATCCCGAGTGGAGTGTGTGAAACACCCCGCCCTTGGTGGAGCTGTGTGTACAGC	1957
Db	1734	CTTGGATCCCGAGTGGAGTGTGTGAAACACCCCGCCCTTGGTGGAGCTGTGTGTACAGC	1793
Qy	1958	TGAGAGGAGGAGCCCATCATCGGCGCGAGAGCTTCTACGTGGAGCGGCGCCCAACCGCG	2017
Db	1794	TGAGAGGAGGAGCCCATCGTGGGCGCGAGAGCTTCTACGTGGAGCGGCGCCCAACCGCG	1853



```
Qy 1358 GGATGGGCTACGAGTGCACCCGACCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417
Db 1194 CCAT-----CGAGCTGCACCCGACCAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1247

Qy 1418 AGGAGAGCTGGACCGTGCACGACATCCAGAGCTGTTGGGCAAGCTGAACCTGGGCGCAGCC 1477
Db 1248 AGGACAGCTGGACCGTGCACGACATCCAGAGCTGTTGGGCAAGCTGAACCTGGGCGCAGCC 1307

Qy 1478 AGATCTACCCCGGCATCAAGTGGCCGACGCTGTGCAAGCTCTCTCGGGCGGCCCAAGGCC 1537
Db 1308 AGATCTACCGCGGCATCAAGTGAAGCAGCTGTGCAAGCTGTCTGGCGGCACCAAGGCC 1367

Qy 1538 TGACCGACATCGTGCCTTGACCGAGAGGCGGAGCTGGAGCTGGCGGAGAACCGCGAGA 1597
Db 1368 TGACCGAGGTGATCCCTTGACCGAGAGGCGGAGCTGGAGCTGGCGGAGAACCGCGAGA 1427

Qy 1598 TCCTGCGGAGCGCGTGCACGCGCTGTACTACGACCCCGACGACCTGGTGGCGGAGA 1657
Db 1428 TCCTGAAGGAGCGCGTGCACGAGGTGTACTACGACCCCGACGACCTGGTGGCGGAGA 1487

Qy 1658 TCCAGAGAGCGGCGCACGACGAGTGGACCTTACAGATCTACGAGAGCCCTTCAAGAAC 1717
Db 1488 TCCAGAGAGCGGCGCAGGCGCAGTGGACCTTACAGATCTACGAGAGCCCTTCAAGAAC 1547

Qy 1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCAACCAACGACGTGAAGCAGCTGA 1777
Db 1548 TGAAGACCGGCAAGTACGCCCGCATGCGCGCGCCACACCAACGACGTGAAGCAGCTGA 1607

Qy 1778 CCGAGGCGGTGCAGAAATCGCCATGAGAGAGCATGTGTATCTGGGGCAAGCCCCAAGT 1837
Db 1608 CCGAGGCGGTGCAGAAAGTGAACCCGAGAGCATCGTGTATCTGGGGCAAGATCCCCAAGT 1667

Qy 1838 TCCGCTCTGCCCATCCAGAGGAGACCTGGGAGACCTGTGTGACCGACTACTTGGCGAGGCCA 1897
Db 1668 TCAAGCTGCCCATCCAGAGAGAGACCTTGGGAGGCTGTGTGATCGAGTACTTGGCGAGGCCA 1727

Qy 1898 CTTGGATCCCGAGTGGAGGTTCGTGAACACCCCGCCCTCTGGTGAAGCTGTGTACAGC 1957
Db 1728 CTTGGATCCCGAGTGGAGGTTCGTGAACACCCCGCCCTCTGGTGAAGCTGTGTACAGC 1787

Qy 1958 TGGAGAGAGGCCCATCATCGGCGCCGAGACCTTCTAGTGGAGCGGCGCGCCCAACCGCG 2017
Db 1788 TGGAGAGAGGCCCATCATCGTGGCGCCGAGACCTTCTAGTGGAGCGGCGCGCCCAACCGCG 1847

Qy 2018 AGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGCGCAGAAATCGTGAGCC 2077
Db 1848 AGACCAAGCTGGGCAAGCGCGCTACGTGACCGACCGGGCGCGCAGAAATCGTGAGCA 1907

Qy 2078 TGACCGGAGACCAACCAAGAGACCGAGCTGCAGGCCATCTCAGCTGGCCCTCGCAGGACA 2137
Db 1908 TCGCCGACACCAACCAAGAGACCGAGCTGCAGGCCATCTCAGCTGGCCCTCGCAGGACA 1967

Qy 2138 CGGCGAGCGAGGTGAACATCGTGAACGACGACGCTGCAGGCCATCTCAGCGCCATCTCAGGCGCC 2197
Db 1968 CGGCGCTGGAGGTGAACATCGTGAACGACGACGCTGCAGGCCATCTCAGCGCCATCTCAGGCGCC 2027

Qy 2198 AGCCCGACAGAGCGAGCGAGCTGTGTGACAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db 2028 AGCCCGACAGAGCGAGCGAGCTGTGTGACAGATCATCGAGCAGCTGATCAAGAAGG 2087

Qy 2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCACAGAGGATCGGGCGCAACAGCAGATCG 2317
Db 2088 AGAAGGTGTACCTGGCTGGGTGCCCGCCACAGAGGATCGGGCGCAACAGCAGGTGG 2147

Qy 2318 ACAAGCTGGTGAACAGGCGATCCGCAAGTGTGTCTTCTGGAACGCGATCGATGGCGGCA 2377
Db 2148 ACAAGCTGGTGAACAGGCGATCCGCAAGTGTGTCTTCTGGAACGCGATCGATGGCGGCA 2207

Qy 2378 TCGTGAATACCACTACAGTACAGACCTGTACGTGGGCGGCGGCGCCCTAGGATCGATT 2437
Db 2208 TCGTGAATACCACTACAGTACAGACCTGTACGTGGGCGGCGGCGCCCTAGGATCGATT 2267
```

```
Qy 2438 AAAAGCTTCCGGGGCTAGCACCCGCTGAATTC 2469
Db 2268 AAAAGCTTCCGGGGCTAGCACCCGCTGAATTC 2299

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

Query Match 68.0%; Score 1678.6; DB 4; Length 4307;
Best Local Similarity 82.9%; Pred. No. 2.8e-278;
Matches 1968; Conservative 0; Mismatches 389; Indels 16; Gaps 4;

Qy 12 CATGCCGAGGCCATGAGCAGG---CCACAGGCGCCAAATCTCTGATGACGCGCAGCAA 68
Db 1086 CTTGGCTGAGGCCATGAGCAGGTCACCACTCCGCTACCATCATGATGACGCGCGCAA 1145

Qy 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGGAGGCCACATCGC 128
Db 1146 CTTTCGGAACAACAGCAAGATCGTCAAGTCTTCAACTGTGGCAAGAGGGCACACAGC 1205

Qy 129 CCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGGCAAGGAGGCCACCA 188
Db 1206 CCGCAACTGCGAGGCCCTTAGGAAAGAGGGCTGTGGAATGCGGCAAGGAGGCCACCA 1265

Qy 189 GATGAAGACTGCACCGAGCGCAGGCGCAACTTTCTTCCGCGAGGACCTTGGCTTCCGCCA 248
Db 1266 GATGAAGACTGTACGAGAGACAGGCTAA--TTTTTTAGGGAAGATCTGGCTTCTCTACA 1324

Qy 249 GGGCAAGGCGCGAGTTCCTCCAGCAGCAGAACCCGCGCCAAACAGCCCAACAGCCCGCA 308
Db 1325 AGGGAAGGCCAGGGAAATTTTCTTCAGAGCAGACCCCGCCCAACAGCCCAACAGCCCGCA 1384

Qy 309 GCTGCAGGT-----GGCGCGCGACAAACCCCGCAGCGAGGCGCGCGCGAGGCCAGGG 362
Db 1385 GCTTCAGTCTGGGGTTCGACAACTCTCCCTTCGGAAGCAGGAGCGCGAGGCCAGGG 1444

Qy 363 CA-----CCCTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCTCTGTGTAGCATCAA 416
Db 1445 CACGGTCTCTCAACTTCCCTCAGGTACGCTTGGCAGCGACCCCTCTGTCTACCATCAA 1504

Qy 417 GGTGGGCGCGCAGATCAAGGAGGCCCTGTGGACACCGCGCGCCAGCACACCGTGTGGA 476
Db 1505 GATCGGGGGGCGAGCTCAAGGAGGCTCTCTTGGACACCGCGAGCAGACACCGTGTGGA 1564

Qy 477 GGAGATGAGCTGCGCCCGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536
Db 1565 GGAGATGCTGTTGCCAGCGCTGGGAAGCCGAGATGATCGGGGAATCGGCGGTTTCAT 1624

Qy 537 CAAGGTGCGCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCA 596
Db 1625 CAAGGTGCGCAGTATGACCAAGATCTCTCATCGAAATCTGCGGGCCCAACAGGCTATCGGTAC 1684

Qy 597 CGTGTCTATCGGCCCCACCCCGTGAACATCATCGGCGCCCAACATGTGTGACCCAGCTGG 656
Db 1685 CGTGTCTATCGGCCCCACCCCGTGAACATCATCGGCGCCCAACATGTGTGACCCAGATCGG 1744
```



QY 657 CTGACCCCTGAATTCCCTCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCG 716  
 Db 1745 TTGCACGCTGAACTTCCCTATAGCCCTATCGAGACGTTACCGTGAAGCTGAAGCCCG 1804  
 QY 717 CATGGACGGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAGATCAAGCCCTGAC 776  
 Db 1805 GATGGACGGCCCAAGGTGAAGCAATGGCCATTGACAGAGGAGAAGATCAAGGCACTGGT 1864  
 QY 777 CGCCATCTGGCAGAGATGAGAGAGGAGGCGCAAGATCAACAAATCGGCCCCGAGAACCC 836  
 Db 1865 GAGATTTGACAGAGATGAGAGAGGAGGAGGAGGAAATCTCCAGATTTGGGCTTGAGAACCC 1924  
 QY 837 CTACAAACACCCCGTGTTCGCCATCAAGAGAGAGGAGCAGCAACAAAGTGGGCGCAAGCTGGT 896  
 Db 1925 GTACAAACACCCCGTGTTCGCCATCAAGAGAGAGGAGCTCAGCAGAAATGGGCGAAGCTGGT 1984  
 QY 897 GGAATTTCCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCAATCC 956  
 Db 1985 GGAATTTCCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTTCAGCTGGGCAATCC 2044  
 QY 957 CCACCCGCGCGCTGAAGAGAGAGAGAGCGTGACCGTCTGAGCGTGGGCGGACGCTA 1016  
 Db 2045 GCACCCGCGAGGCTGAAGAGAGAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCA 2104  
 QY 1017 CTTTCAGCGTCCCTTGACGAGGACTTCCGCAAGTACACCGCTTACACATCCCCAGCAT 1076  
 Db 2105 CTTTCAGCGTCCCTTGACGAGGACTTCCGCAAGTACACCGCTTACACATCCCTTCGAT 2164  
 QY 1077 CAACAAAGAGACCCCGGCTCCGCTACCAAGTCAACAGTCTGCCCGAGGCTGGAAGG 1136  
 Db 2165 CAACAAAGAGACCCCGGCTCCGCTACCAAGTCAACAGTCTGCCCGAGGCTGGAAGG 2224  
 QY 1137 CAGCCCGAGCATTTCCAGAGAGCATGACCAAGATCCTGGAGCCCTTCGCGCCGCGCA 1196  
 Db 2225 CTTCTCCGCAATTTCCAGAGTACATGACCAAAATCCTGGAGCCCTTCGCGCAAAAG 2284  
 QY 1197 CCCGAGATCGTATCTACAGTACATGAGAGACTGTAGTGGGAGCGACCTTCGCGCCGCGCA 1256  
 Db 2285 CCCGAGATCGTATCTATCAGTACATGATGATGATGATGATGATGATGATGATGATGAT 2344  
 QY 1257 CGGCCAGCACCGGCCCAAGATCAGAGAGTGCACAAAGACCTGCTGGCTGGGCTTCAC 1316  
 Db 2345 AGGCGAGCACCGGCCCAAGATCAGAGAGTGCACAAAGACCTGCTGGCTGGGCTTCAC 2404  
 QY 1317 CACCCCGCAAGAGCACAG 1376  
 Db 2405 CACCCCGCAAGAGCACAG 2464  
 QY 1377 CCGCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGAGAGAGAGAGAGAG 1436  
 Db 2465 CCGCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGAGAGAGAGAGAGAG 2524  
 QY 1437 CGACATCCAGAGCTGGTGGGCAAGTGAATGGGCGAGCCAGATCTACCCCGGCTCA 1496  
 Db 2525 CGACATACAGAGCTGGTGGGCAAGTGAATGGGCGAGCTGAGTGGGCTGAGTGGGCTGAGT 2584  
 QY 1497 GGTGGCCAGCTGTGACAGTGTGCGCGGCGGAGGCGCTACCGACATCTGCGCCCT 1556  
 Db 2585 GGTGGCCAGCTGTGACAGTGTGCGCGGCGGAGGCGCTACCGACATCTGCGCCCT 2644  
 QY 1557 GACCGAGGAGCGGAGCTGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1616  
 Db 2645 AACCGAGGAGCGGAGCTGAACTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2704  
 QY 1617 CGCGGTGTACTACAGCCCGAG 1676  
 Db 2705 CGCGGTGTACTACAGCCCGAG 2764  
 QY 1677 CAGTGGACCTACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1736  
 Db 2765 CCAGTGGACCTACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2824

QY 1737 CAAAGATGCGCACCGCCACACACAAACGAGCGTGAAGCAGCTGACCGAGCCCGTGCAGAGAT 1796  
 Db 2825 CCGGATGAGGGTGCCCAACACACAAACGAGCTGAAGCAGCTGACCGAGCCCGTGCAGAGAT 2884  
 QY 1797 CGCCATGAGAGAGATCGTGAATCTGGGGCAAGACCCCAAGTTCGGCTGCCCATCCAGAA 1856  
 Db 2885 CACCAACGAGAGATCGTGAATCTGGGGCAAGACCTCTAAGTTCAAAGTTCGCCCATCCAGAA 2944  
 QY 1857 GAGAGCTGGAGAGACCTGGTGGACCGACTACTGGCAGGCGCACCTGGATCCCGAGTGGGA 1916  
 Db 2945 GGAACCTGGGAACCTGGTGGACAGATATTGGCAGGCGCACCTGGATTCCTGAGTGGGA 3004  
 QY 1917 GTTTCGTAACACACCCCGCTGGTGAAGCTGGTACCAAGTGGAGAGAGAGAGAGAGATCAT 1976  
 Db 3005 GTTTCGTAACACACCCCGCTGGTGAAGCTGGTACCAAGTGGAGAGAGAGAGAGATCAT 3064  
 QY 1977 CGCGCGGAGACCTTCTAGCTGGACCGCGCGCCCAACCGCGAGACCAAGATCCGCAAGGC 2036  
 Db 3065 GCGCGCGGAGACCTTCTAGCTGGATGGGGCGCTTAACAGGGAGACTAAGCTGGGCAAGC 3124  
 QY 2037 CGCTACGTGACCGACCGCGCGCGCGAGAGATCGTGAGCTGACCGAGACCAACCAACCA 2096  
 Db 3125 CGGATACGTCACTAACCGGGCGAGACAGAGGTTGTCAACCTCAGTGCACACCAACCA 3184  
 QY 2097 GAAGACCGAGCTGACAGGCCCATCAGCTGGCCCTTCAGGAGCAGCGGAGAGAGAGAGAT 2156  
 Db 3185 GAAGACCGAGCTGACAGGCCCATCAGCTGGCCCTTCAGGAGCTCGGGCTGGAGTGAACAT 3244  
 QY 2157 CGTGACCGAGCGCAGTACGCCCTGGGCATCATCAGGCGCCAGCCCGAGAGAGAGAGAG 2216  
 Db 3245 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTCAGGCGCCAGCCAGAGAGAGAGATC 3304  
 QY 2217 CGAGCTGTTGAACACAGATCATCGAGCAGCTGATCAAGAGAGAGAGAGAGAGTGTAC 2276  
 Db 3305 CGAGCTGTTGAACACAGATCATCGAGCAGCTGATCAAGAGAGAGAGAGAGTGTATCTG 3364  
 QY 2277 GGTGCGCCCGCCCAAGGGCATCGCGGCAACCGAGCAGATCGACAAAGCTGGTGAAGAGG 2336  
 Db 3365 GGTGCGCCCGCCCAAGGGCATCGCGGCAATGAGCAGGTCGACAAAGCTGGTCTCGGCTG 3424  
 QY 2337 CATCCGAGGTGCTGCTTCTGACCGGCATCGA 2369  
 Db 3425 CATCAGGAAGTGTATTCTTGGATGGCATCGA 3457

RESULT 8  
 ; US-09-936-572-2  
 ; Sequence 2, Application US/09936572  
 ; Patent No. 6783981  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UDEN, MARK  
 ; APPLICANT: MITROPHANOUS, KYRIACOS  
 ; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
 ; FILE REFERENCE: 078883/0137  
 ; CURRENT APPLICATION NUMBER: US/09/936,572  
 ; CURRENT FILING DATE: 2001-12-11  
 ; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
 ; PRIOR FILING DATE: 2000-03-17  
 ; PRIOR APPLICATION NUMBER: GB 9906177.2  
 ; PRIOR FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 4307  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:  
 ; OTHER INFORMATION: gagpol-SYNGp-codon optimised gagpol sequence  
 ; US-09-936-572-2

Query Match 67.2%; Score 1659.4; DB 4; Length 4307;  
 Best Local Similarity 82.4%; Pred. No. 5.3e-275;

		Matches 1956; Conservative 0; Mismatches 401; Indels 16; Gaps 4;			
Qy	12	CATGCCGAGCCATGAGCCAGG---CCACAGGCGCAACATCTCTGATGACGCGCAGCAA	68		
Db	1086	CCTGGCTGAGGCCATGAGCCAGGTTGACCACTCCGCTACCATCATGATGACGCGCGCAA	1145		
Qy	69	CTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGAGGCCACATCGC	128		
Db	1146	CTTTCGGAACAACGCAAGATCGTCAAGTGTCTTCAACTGTGTGCAAGAAGGGCACAGC	1205		
Qy	129	CCGCAACTGCCCGCCGCCCAAGAAAGGGGTCTGGAAGTGCAGGCGCAAGGAGGCCACCA	188		
Db	1206	CCGCAACTGCGAGGCCCTTAGGAAGAGGGGTCTGGAATGCGGCAAGGAGGCCACCA	1255		
Qy	189	GATGAAGCATGTCACGAGGCGCCAGGCCAACTTCTTCGCGAGCACTTGGCCTTCCGCCA	248		
Db	1266	GATGAAGCATGTACTGAGAGACAGGGCTAA-TTTTTTAGGGAAGATCTGGCCTTCTCTACA	1324		
Qy	249	GGCAAGGCCCGGAGTTCCCGAGGAGCAGAAACCGCGCCCAACAGCCGCCACCGCCGGA	308		
Db	1325	AGGAAGGCCAGGGAAATTTTCTTCAGAGCAGACAGAGCCCAACAGCCGCCACCGAAGAGA	1384		
Qy	309	GCTGACGGTGGGGTAGAGNACAACACTCCCTCTCAGAGCAGAGCGCGATAGCAAGG	1444		
Db	1385	GCTTCAGGTCTGGGGTAGAGNACAACACTCCCTCTCAGAGCAGAGCGCGATAGCAAGG	1444		
Qy	363	CA-----CCCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCTCTGTGAGCATCAA	416		
Db	1445	AACGTGATCTTTAACTTCCCTCAGATCACTCTTTTGGCAACGACCCCTCGTCACAATAAA	1504		
Qy	417	GGTGGGCGCCAGATCAAGAGGCGCTGTGGACACCGGCGCCGACGACACCGTGTGGGA	476		
Db	1505	GATAGGGGGGAGCTCAAGGAGGCTCTCTGGACACCGGAGCAGACGACACCGTGTGGGA	1564		
Qy	477	GGAGATGAGCTGCCCGCAAGTGGAGCCCAAGATGATCGGCGGCATCGGCGCTTCAT	536		
Db	1565	GGAGATGAGCTGCCCGCAAGTGGAGCCCAAGATGATCGGCGGCATCGGCGCTTCAT	1624		
Qy	537	CAAGGTGCGCCAGTACGACAGATCTCTGATCGAGATCTGGGCAAGAGGCCATCGGCGAC	596		
Db	1625	CAAGGTGCGCCAGTATGACAGATCTCTCAGATCTGCGGCAAGAGGCCATCGGCGAC	1684		
Qy	597	CGTGTGATCGGCCCGCCACCCCGTGAACATCATCGGCGCGCAACATGTGACCCAGCTGGG	656		
Db	1685	CGTGTGATCGGCCCGCCACCCCGTGAACATCATCGGCGCGCAACATGTGACCCAGCTGGG	1744		
Qy	657	CTGACCCCTGAACTTCCCGCATCAGCCCATCGAGCCGTCGCTGAAGCTGAGCCCGG	716		
Db	1745	TTGACCGCTGAATTTCCCGCATCAGCCCATCGAGCCGTCGCTGAAGCTGAGCCCGG	1804		
Qy	717	CATGAGCGGCCCGCCAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGAC	776		
Db	1805	GATGAGCGGCCCGCCAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGAC	1864		
Qy	777	CGCATCTCGAGAGATGAGAGAGGAGGCAAGATCACAGATCGGCGCCCGAGAACCC	836		
Db	1865	GGAGATTTGACAGATGAGAGAGGAGGAGGAAATCTCCAAGATTGGGCGCTGAGAACCC	1924		
Qy	837	CTAACACACCCCGTGTTCGCAATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGT	896		
Db	1925	GTACACACCGCGGTGTTCGCAATCAAGAGAGGAGCTCGACGAAATTTGGCGCAAGCTGGT	1984		
Qy	897	GGACTTCGCGGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCGAGCTGGGCAATCCC	956		
Db	1985	GGACTTCGCGGAGCTGAACAGCGCACCGAGACTTCTGGAGGTTCAGCTGGGCAATCCC	2044		
Qy	957	CCACCCCGCGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGCTGGGCGCGACCTTA	1016		
Db	2045	GCACCCCGCGCGCTGAAGAGAGAGAGAGCGTGAACCGTGTGAGCTGGGCGCGACCTTA	2104		
Qy	1017	CTTCAGGCTGCCCTGACGAGGACTTCCGCAAGTACACCGCCTTTCACCATCCCGCAT	1076		
Db	2105	CTTCTCCGTTTCCCGTGGAGGAGCTTTCAGGAAGTACACTGCCTTTCACATCCCTTCGAT	2164		

Qy	1077	CAACAAACGAGACCCCCCGGCATCCGCTACCACTAGTACAAAGTGTGCTGCCAGGCTGGAAGG	1136		
Db	2165	CAACAAACGAGACCCCCCGGCATCCGCTACCACTAGTACAAAGTGTGCTGCCAGGCTGGAAGG	2224		
Qy	1137	CAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGCCCTTCGCGCCCGCAA	1196		
Db	2225	CTCTCCGCAATCTTCAGAGTAGCATGACCAAAATCTCGAGCTTTTCCGCAACAGAA	2284		
Qy	1197	CCCCGAGATCGTGTATCTACCACTAGTACATGAGCAGCTGTACGTTGGGCGAGCAGCTGGAGAT	1256		
Db	2285	CCCCGACATCGTCTATCTCAGTACATGATGACTTGTACGTGGGCTCTGATCTAGAGAT	2344		
Qy	1257	CGGCGAGACCGCGCCCAAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGGCTTCAAC	1316		
Db	2345	AGGCGAGACCGCGCCCAAGATCGAGAGCTGCGCAGCACCTGTTGAGGTGGGACTGAC	2404		
Qy	1317	CACCCCGCAGCAAGAACCAACAGAGGAGCCCTCTCTCTGATGAGTGGGCTACGAGCTGCA	1376		
Db	2405	CACCCCGCAGCAAGAACCAACAGAGGAGCCCTCTCTCTGATGAGTGGGCTACGAGCTGCA	2464		
Qy	1377	CCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGCAACCGTGA	1436		
Db	2465	CCCTGACAAATGGACCGTGCAGCTATCTGCTGCCAGAGAAAGACAGCTGGAATGCAA	2524		
Qy	1437	CGACATCCAGAGCTGTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGATCAA	1496		
Db	2525	CGACATCCAGAGCTGTGTGGGCAAGCTGAACTGGGCGCAGCTGAGATTTACCCAGGATTA	2584		
Qy	1497	GGTGGCGCAGCTGTGCAAGCTGTGCGGGCGCAAGCCCTGACCGACATCGTCCCTCT	1556		
Db	2585	GGTGGCGCAGCTGTGCAAGCTGTGCGGGCGCAAGCCCTGACCGACATCGTCCCTCT	2644		
Qy	1557	GACCGAGGAGCGGAGCTGAGCTGGCGGAGAACCGCGAGATCTCTGCGAGCCGCTGCA	1616		
Db	2645	AACCGAGGAGCGGAGCTGAGCTGGCGGAGAACCGCGAGATCTCTAAGGAGCCGCTGCA	2704		
Qy	1617	CGGCGTGTACTACGACCCCGCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGCGCA	1676		
Db	2705	CGGCGTGTACTACGACCCCGCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGCGCA	2764		
Qy	1677	CGAGTGAACCTACAGATCTACCGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTAGCG	1736		
Db	2765	CGAGTGAACCTACAGATCTTACCGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTAGCG	2824		
Qy	1737	CAAGTGCAGACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGAT	1796		
Db	2825	CCGATGAGGGGTGCCCACTAACGACGCTCAGCAGCTGACCGAGGCGGTGCAAGAGAT	2884		
Qy	1797	CGCCATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAA	1856		
Db	2885	CACCAACGAAAGCATCTGTATCTGGGCAAGACCTCTTAAGTTCAAGCTGCCCATCCAGAA	2944		
Qy	1857	GGAGACCTGAGAGACCTGTGTGGACCGACTACTGCGAGCCACCTGGATCCCGGATGGGA	1916		
Db	2945	GGAAACCTGAGAGACCTGTGTGGACCGACTACTGCGAGCCACCTGGATCCCGGATGGGA	3004		
Qy	1917	GTTTCGTGAACACACCCCGCTGTGTGAAGCTGTGTACAGCTGGAAGAGAGGAGCCATCAT	1976		
Db	3005	GTTTCGTGAACACACCCCGCTGTGTGAAGCTGTGTACAGCTGGAAGAGAGGAGCCATCAT	3064		
Qy	1977	CGGCGCGAGACCTTCTACGTGGAACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC	2036		
Db	3065	GGGCGCGCAACCTTCTACGTGGAACCGCGCGCCCTAACAGGAGACTAAGCTGGGCAAGGC	3124		
Qy	2037	CGGCTAGTGAACCGACCGGGCGCGGAGAGATGTTGAGGCTGACCGAGACCAACCAACCA	2096		
Db	3125	CGGATACGTCACTAAACCGGGCGGAGACAGAAAGTTGTCACTCTGACCAACCAACCA	3184		
Qy	2097	GAAGACCGAGCTGCGAGGCCATCCAGCTGGCCCTTGCAGGACAGCGGCGAGGAGTGAACAT	2156		
Db	3185	GAAGACCGAGCTGCGAGGCCATTTACCTCGCTTTTGCAGGACTCGGGCTCGGAGGTGAACAT	3244		

QY 2157 CGTGACCGACGAGCCAGTACGCTGGGCGATCATCCAGGCGCCAGCCGACCAAGAGCGAGAG 2216  
 Db 3245 CGTGACAGACTCTCAGTATGCCCTGGGCATCATTTCAAGCCCGACGACGAGTGGTC 3304  
 QY 2217 CGAGCTGGTGAACACAGATCATCAGACAGCTGATCAAGAGGAGAAAGGTGATCCTGAGCTG 2276  
 Db 3305 CGAGCTGGTCAATCAGATCATCAGACAGCTGATCAAGAGGAGAAAGGTCTATCTGGCCTG 3364  
 QY 2277 GGTGCGCGCCCAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAAGCAAGGG 2336  
 Db 3365 GGTACCGCGCCCAAGAGGCAATGGCGCAATGAGCAGGTGCAACAAGCTGGTCTCGGCTGG 3424  
 QY 2337 CATCCGCAAGGTGCTGTTCTGACCGCATCGA 2369  
 Db 3425 CATCAGAGGTGCTATTCTGGATGGCATCGA 3457

RESULT 9

US-09-936-572-14  
 ; Sequence 14, Application US/09936572  
 ; Patent No. 6783981  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UDEN, MARK  
 ; APPLICANT: MITROPHANOUS, KYRIACOS  
 ; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
 ; FILE REFERENCE: 078983/0137  
 ; CURRENT APPLICATION NUMBER: US/09/936, 572  
 ; CURRENT FILING DATE: 2001-12-11  
 ; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
 ; PRIOR FILING DATE: 2000-03-17  
 ; PRIOR APPLICATION NUMBER: GB 9906177.2  
 ; PRIOR FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: PatentIn ver. 2.1  
 ; SEQ ID NO 14  
 ; LENGTH: 4327  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: pSYNFP4-codon  
 ; OTHER INFORMATION: optimised HIV-1 gagpol with 20bp of the leader  
 ; OTHER INFORMATION: sequence of HIV-1  
 US-09-936-572-14

Query Match 66.9%; Score 1651.4; DB 4; Length 4327;  
 Best Local Similarity 82.2%; Pred. No. 1.2e-273;  
 Matches 1951; Conservative 0; Mismatches 406; Indels 16; Gaps 4;  
 QY 12 CATGGCGGAGCCATGAGCCAGG---CCACGAGCGCCAAACATCTGATGAGCGGCGACAA 68  
 Db 1106 CTGGCTGAGCCATGAGCCAGGTGACCACTCGGCTACCATCATGATGAGCGCGGCAA 1165  
 QY 69 CTTCAAGGGCCCAAGCGCATCATAGTCTTCAACTGCGGCAAGAGGGCCACATTCGC 128  
 Db 1166 CTTTCGGAACCAAGCAAGATGCTCAAGTGTCTTCAACTGTGGCAAGAGGGCACACAGC 1225  
 QY 129 CCCTAACTGCGGCGCCCGCCCGAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCA 198  
 Db 1226 CCGCAACTGCAAGGCGCCCTAGGAAAAGGGCTGTGGAAATGTGGAAGAGGACACCA 1285  
 QY 189 GATGAAGGACTGCACGAGCGCGCAGGCCAACTTCTTCGCGGAGGACCTGGCCCTTCCCCCA 248  
 Db 1286 AATGAAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGAAGATCTGGCCCTTCCACA 1344  
 QY 249 GGGCAAGGCCCGGAGTTCCCGAGCGAGCAAGAACCGCGCCAAACAGGCCCAACCGCGCGA 308  
 Db 1345 AGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACAGAGCCAAACAGCCCCACCAAGAGA 1404  
 QY 309 GCTGCAGGTGCGGG-----CGACACCCCGCAGCGAGCGCGCGCGAGCGCCAGGG 362  
 Db 1405 GCTTCAGGTTTGGGGAGAGACAACTCCCTCTCAGAGAGGAGGCGCATAGACAAGG 1464  
 QY 363 CA-----CCCTGAACTTCCCCCAGATCACCTGTGGCAGCGGCCCTGTGAGCATCAA 416

Db 1465 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAATAAA 1524  
 QY 417 GGTGGGGGGCGAGATCAAGAGGCGCTGCTGGACACCGCGCGCGAGACACCGTGTGGA 476  
 Db 1525 GATAGGGGGGCGAGCTCAAGGAGGCTCTCTGGACACCGGAGCAGACACCGTGTGGA 1584  
 QY 477 GGAGATGAGCTGCGCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCAT 536  
 Db 1585 GGAGATGTGTTGGCAGGCGCTGGGAAGCCGAGATGATCGGGGAATCGGCGGTTTCAT 1644  
 QY 537 CAAAGTGCGCCATACGACAGATCCTGATCGAGATCTCGGCGCAAGAGGCCATTCGGCAC 596  
 Db 1645 CAAAGTGCGCCATGATGACAGATCCTCATCGAAATCTCGGCGCAAGAGCTATCGGTAC 1704  
 QY 597 CGTGTGATCGGCGCCCGCCCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGG 656  
 Db 1705 CGTGTGTTGGGCGCCCGCACACCCCGTCAACATCATCGGACGCACTGTGTGACGAGATCGG 1764  
 QY 657 CTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGG 716  
 Db 1765 TTGCACGCTGAACTTCCCATTTAGCCCTATCGAGACGCTACCGTGAAGCTGAAGCCCGG 1824  
 QY 717 CATGACCGGCCCAAGGTGAAGCAGTGGCCCTTGAACGAGGAGAAAGATCAAGGCCCTGAC 776  
 Db 1825 GATGGACGCGCCCGAAGGTCAAGCAATGGGCCATTGACAGAGGAGAAAGATCAAGGCACTGGT 1884  
 QY 777 CGCATCTCGGAGGAGATGGAGNAGGAGGCGCAGATCAACAGATCGGCCCGCGAGACCC 836  
 Db 1885 GGAGATTTTGCACAGAGATGAAAGAGGAGAAATCTCAAGATTTGGGCTCGAGAACCC 1944  
 QY 837 CTACAAACACCCCGCTGTTCGCCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGGT 896  
 Db 1945 GTACACACGCGCGTGTTCGCCATCAAGAGAGAGGACTCGACGAATGGCGCAAGCTGGT 2004  
 QY 897 GGACTTTCGCGAGCTGAAACAGCGCAACAGCGCATCTTCTGGGAGGTGAGCTGGGCGATCCC 956  
 Db 2005 GGACTTTCGCGAGCTGAAACAGCGCACGCAAGACTTCTGGGAGGTTTCAGCTGGGCGATCCC 2064  
 QY 957 CCACCCCGCGGCTCAAGAGAGAGAGCGCTGACCGCTGCGAGTGGGCGCAGCGCTA 1016  
 Db 2065 GCACCCCGCAGGGCTGAAGAAGAGAAATCCCGTACCGTACTGATGTGGGTGATGCGCTA 2124  
 QY 1017 CTTACGCTGCCCTCGAGAGGACTTTCGCAAGTACACCGCTTCCACATCCCGCAGCAT 1076  
 Db 2125 CTTCTCGTTCCTCGACGAGACTTTCAGAGAGTACACTGCTTCACAATCCCTTCGAT 2184  
 QY 1077 CAACACGAGACCCCGGCAATCCGCTACCAAGTACAAAGCTGCTGCCCGCAGGGCTGGAAGGG 1136  
 Db 2185 CAACACGAGACACCGGGGATTCGATATCAGTACAAAGCTGCTGCCCGCAGGGCTGGAAGG 2244  
 QY 1137 CAGCCCGCAGATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCGCAA 1196  
 Db 2245 CTTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTTGGAGCCTTTCGCGAAACAGAA 2304  
 QY 1197 CCGCGAGATCGTCTATCTACCAAGTACATGAGACCTGTACGTGGGCGAGCAGCAGCTGGAGAT 1256  
 Db 2305 CCGCGACATCGTCTATCTTACGATACATGATGATCTGTAGTGGGCTCTGATCTAGAGAT 2364  
 QY 1257 CGGCCAGCACCGCGCCCAAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGGCTTCAC 1316  
 Db 2365 AGGGCAGCACCGCACCAAGATCGAGAGCTGCGCCAGCACCTGTGTAGGTGGGAGCTGAC 2424  
 QY 1317 CACCCCGCAGCAAGACCAAGAGGAGGCGCCCTTCTGTGATGGGCTACGAGCTGCA 1376  
 Db 2425 CACACCCGCAAGAGCAACAGAGGAGGCGCTCCCTTCTCTGTGATGGGTACGAGCTGCA 2484  
 QY 1377 CCGCGCAAGTGAACCGTGCAGGCCATCGAGCTGCCCGAGAGGAGAGCTGGACCGTGAA 1436  
 Db 2485 CCCTGACAAATGACCGTGCAGCTTATCGTGTCTGCCAGAGAGAGACAGCTGGAGTCAA 2544  
 QY 1437 CGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAA 1496

Db 2545 CGACATACAGAAAGCTGGTGGGAAGTTGAACCTGGCCAGTCAGATTTACCAGAGGATTAA 2604  
Qy 1497 GGTGCGCAGCTGTCAAGCTGCTCGCGGGCCCAAGGCCCTTGACCGACATCGTGCCTT 1556  
Db 2605 GGTGAGGCGAGCTGTGCAAACTCTCCGCGGAACCAAGGCACTCAAGAGGTGATCCCTT 2664  
Qy 1557 GACCGAGAGCCGAGCTGGAGCTGGCCGGAACCGCGAGATCTTGGCGAGCCGCTGCA 1616  
Db 2665 AACCGAGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCTTAAAGGAGCCGCTGCA 2724  
Qy 1617 CGGCTGTACTACGACCCACGACAGGACCTGGTGGCCGAGATCCAGAAAGCAGGCGCCACGA 1676  
Db 2725 CGGCTGTACTATGACCCCTCAAGACCTGATCGCCGAGATCCAGAAAGGCGGCAAG 2784  
Qy 1677 CCAGTGGACCTACAGATCTACAGAGGACCTTCAAGAACTGAAAGCCGCAAGTAGTC 1736  
Db 2785 CCAGTGGACCTACAGATTTACAGAGGACCTTCAAGAACTGAAAGCCGCAAGTAGTC 2844  
Qy 1737 CAAGATGGCACCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAT 1796  
Db 2845 CCGATGAGGGGTGCCCACTAAACGAGCTCAAGCAGCTGACCGAGGCGGTGCAAGAT 2904  
Qy 1797 CGCATGAGAGCATCTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAA 1856  
Db 2905 CACCAACGAAAGCATCTGATCTGGGAAAGACTCTTAAGTTCAAGCTGCCCATCCAGAA 2964  
Qy 1857 GGAGACCTGGGAGACCTGGTGGACCGACTATGGCAGGCCACCTGGATCCCGAGTGGGA 1916  
Db 2965 GGAAACCTGGGAAACCTGGTGGACAGAGTATGGCAGGCCACCTGGATTCCTGAGTGGGA 3024  
Qy 1917 GTTCTGTAAACCCCCCTGGTGAAGCTGTGTACCACTGGAGAGGAGGCCATCAT 1976  
Db 3025 GTTCTGTAAACCCCCCTGGTGAAGCTGTGTATCCAGCTGGAGAGGAGGCCATCAT 3084  
Qy 1977 CGGCGCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2036  
Db 3085 GGGCGCGGAAACCTTCTACGTGGATGGGCGCGCTTAACAGGAGACTAAGCTGGGCAAGC 3144  
Qy 2037 CGGTACGTGACCGACCGGGCCGCGCAGAAAGATCGTGAGCCTGACCGAGACCAACCA 2096  
Db 3145 CGGATACGTCACTAACCGGGCCAGACAGAGGTTGTCAACCTCACTGACACCAACCA 3204  
Qy 2097 GAAGACGAGCTGAGGCGCATCAGCTGGCCCTCGCAGGACGAGCGGAGGTGAACAT 2156  
Db 3205 GAAGACTGAGCTGAGGCGCATTTACCTCGCTTTGACAGACTCGGGCCCTGGAGGTGAACAT 3264  
Qy 2157 CGTGACGACGAGCAGTACGCTGGGCATCATCCAGGCCAGCCCGACAGAGCGGAG 2216  
Db 3265 CGTGACAGACTCTCAGTATGCCCTTGGGCATCATTTCAAGCCCGACAGACAGGTGAGTC 3324  
Qy 2217 CGAGCTGGTAAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTG 2276  
Db 3325 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAGGAAAGGTCTATCTGGCCTG 3384  
Qy 2277 GGTGCGCCGCCAAGGGCATCGCGGCAACGAGCAGATGCAAGCTGTGTGACAAAGG 2336  
Db 3385 GGTACCGCCGCCAAGAGCATTTGGCGCAATGAGCAGGTGCAAGCTGTCTCGGCTGG 3444  
Qy 2337 CATCGCAAGGTGCTGTTCTGGAGCGGCATCGA 2369  
Db 3445 CATCAGAAAGTGCTATTCTTGGATGGCATCGA 3477

## RESULT 10

US-09-936-572-13

; Sequence 13, Application US/09936572

; Patent No. 6783981

; GENERAL INFORMATION:

; APPLICANT: UDEN, MARK

; APPLICANT: MITROPHANOUS, KYRIACOS

; TITLE OF INVENTION: ANTI-VIRAL VECTORS

; FILE REFERENCE: 078863/0137

; CURRENT APPLICATION NUMBER: US/09/936,572

; CURRENT FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: GB 9906177.2  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 4353  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PSYNGP3-codon  
; OTHER INFORMATION: optimised HIV-1 gapcol with leader sequence from  
; OTHER INFORMATION: the major splice donor  
US-09-936-572-13

Query Match 66.9%; Score 1651.4; DB 4; Length 4353;

Best Local Similarity 82.2%; Pred. No. 1.2e-273;

Matches 1951; Conservative 0; Mismatches 406; Indels 16; Gaps 4;

Qy 12 CATGCCGAGGCCATGAGCCAGG---CCACCAGCGCCCAACATCTGTATGTCAGCGCAGCAA 68  
Db 1132 CTTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGAGCGGGCAA 1191  
Qy 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGGAGGCCACATCGC 128  
Db 1192 CTTTCGGAAACAACCGCAAGATCGTCAAGTCTTCAACTGTGCAAAAGAGGCCACACAGC 1251  
Qy 129 CCGCAACTGCGCGCCCGCCCGCAGAGAGGCTGTGGAAGTCCGCGAGGAGGGCCACCA 188  
Db 1252 CCGCAACTGCGAGGCCCCCTAGGAAAAAGGGCTGTGGAATGTGGAAGGAGGACCA 1311  
Qy 189 GATGAAGACTGCAACCGAGCGCCAGGCAACTTCTTCGCGAGGACCTGGGCTTCCGCCA 248  
Db 1312 AATGAAGATTTGACTGAGAGACAGGCTAA-TTTTTAGGAAGATCTGGCTTCCCA 1370  
Qy 249 GGGCAAGGCCCGGAGTTCCCGCAGCAGAGAGAAACCGCGCCAAACAGCCCAACAGCGCGA 308  
Db 1371 AGGAAGGCCAGGGAATTTTCTTCAGAGCAGACAGAGCCAAACAGCCCAACAGAGAGA 1430  
Qy 309 GCTCAGGTGGCGG-----CGACAAACCCCGCAGCGAGCGCGCGCGAGCCCGAGG 362  
Db 1431 GCTTCAGGTTTGGGGAAGAGACAACTCCCTCTCAGAGCAGGAGCCGATAGCAAGG 1490  
Qy 363 CA-----CCCTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCTCTGGTGAGCATCAA 416  
Db 1491 AACTGTATCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCTGTCANATAA 1550  
Qy 417 GGTGGCGGCAGATCAAGGAGGCCCTGCTGGACACCCGCGCCGACGACACCGTGTGGA 476  
Db 1551 GATAGGGGGCAGCTCAAGGAGGCTCTCTGGACACCGGAGCAGACACCGTGTGGA 1610  
Qy 477 GGAGATGAGCTGCGCGGAAAGTGGAAAGCCCAAGATGATCGGCGCATCGGCGGCTTCAT 536  
Db 1611 GGAGATTCGTGTGCGAGCGCGCTGGAAAGCCGAAGTATCGGGGGAATCGGCGGTTTCAT 1670  
Qy 537 CAAGTGGCGCAGTACCAACAGATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAC 596  
Db 1671 CAAGTGGCGCAGTATGACAGATCTCTCATCGAATCTGCGGCCCAAGGCTATCGGTAC 1730  
Qy 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGG 656  
Db 1731 CGTGTGTTGGGGCCACACCCCGTCAACATCATCGGACGCAACCTGTTGACCGAGATCGG 1790  
Qy 657 CTGCACCTGAACTTCCCGATCAGCCCGCATCGAGCCGTGCGCCGTGAAGTGAAGCCGG 716  
Db 1791 TTGCACGCTGAACCTTCCCGATTAGCCCTATCGAGACGTTACCGGTGAAGCTGAAGCCGG 1850  
Qy 717 CATGAGCGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGAGAGAGTCAAGGCGCTGAC 776  
Db 1851 GATGAGCGGCCCGAAGGTCAAGCAATGGCCATTTGACAGAGAGAGAGATCAAGGCACTGGT 1910

Qy	777	CGCCATCTGCGAGAGATGGAAGAAGGAGGCAAGATCACCAAGATCGGCCCGCCGAGAACCC	836
Db	1911	GGAGATTTGCAACAGATGGAAGAAGGAGGAAAAATCTCCAAGATTGGGCTCGAGAACCC	1970
Qy	837	CTACAAACACCCCGTGTTCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCGCAAGCTGCT	896
Db	1971	GTACAAACACGCCGGTGTTCGCAATCAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGT	2030
Qy	897	GGACTTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCC	956
Db	2031	GGACTTCCGCGAGCTGAACCAAGCGCACCGCAAGACTTCTGGGAGGTTTCAGCTGGGCATCCC	2090
Qy	957	CCACCCCGCGCTGAAGAAGAAAGAGAGCGTGACCGTCTGGAACGTGGCGGACGCGCTA	1016
Db	2091	GCACCCCGCAGGGCTGAAGAAGAAGAAATCCGTGACCCGTATCGATGTGGTGTATGCGCTA	2150
Qy	1017	CTTCAGGTGCGCCTGGAACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCCAGCAT	1076
Db	2151	CTTCTCGCTTCCCCTGGAACGAAGACTTCAAGAAAGTACACTGCGCTTCACAATCCCTTCGAT	2210
Qy	1077	CAACAAACGAGACCCCGGCATTCGCTACCAAGTACAACGCTGCTGCCAGGCGCTGGAAGGG	1136
Db	2211	CAACAAACGAGACCCGGGATTCGATATCAGTACAGCTGCTGCCAGGCGCTGGAAGG	2270
Qy	1137	CAGCCCCAGCATCTTTCAGAGCAGCATGACAAGATCCTGGAGCCCTTCGCGCGCCGCAA	1196
Db	2271	CTCTCCCGCAATCTTTCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCGCGCAACAGAA	2330
Qy	1197	CCCCGAGATCGTGTATCTACCAATGATGAGGACCTGTACGTGGGAGCGACCTGGAGAT	1256
Db	2331	CCCCGAGATCGTCACTCTATCAGTACATGGAATGACTTGTACGTGGGCTCTGATCTAGAGAT	2390
Qy	1257	CGGCGACACCGGGCCAAAGATCGAGGAGCTGCGCAAGCACTGCTGGCTGGGGCTTCAC	1316
Db	2391	AGGCGACACCGCACCAAGATCGAGGAGCTGCGGCAGACACTGTTGAGTGGGAGCTGAC	2450
Qy	1317	CACCCCGACAAGAAGACCAAGAGGAGCGCCCTCTCTGTGTGATGGGCTACGAGCTGCA	1376
Db	2451	CACACCCGACAAGAAGCACCAAGAGGAGCTCTCCCTCTCTGGATGGTTTACGAGCTGCA	2510
Qy	1377	CCCCGACAAGTGACCGTGACGCCATCGAGCTGCTGCCGAGAGAGAGCTGGACCGTGAA	1436
Db	2511	CCCTGCAAAATGGAACCGTGAGCCTATCTGTGCTGCCAGAGAAAGACAGCTGGACTGTCAA	2570
Qy	1437	CGACATCCAGAAGCTGTGGGCGAAGCTGAACATGGGCGCAGCCAGATCTACCCCGGCATCAA	1496
Db	2571	CGACATACAGAAGCTGTGGGGAAATGAACTGGGCCAGTCTGAGATTTACCACGGGATTA	2630
Qy	1497	GGTGCGGCAGCTGTGCAAGCTGTGCGGGCGCCAGGCGCCCTGACCGACATCGTGCCCT	1556
Db	2631	GGTGAGGCAGCTGTGCAAACTCTCTCCGCGGAAACCAAGGCACCTCACAGAGTGTATCCCT	2690
Qy	1557	GACCGAGGAGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGGCGAGCCCGTGCA	1616
Db	2691	AAACCGAGGAGGCGAGCTCGAACTGGCGAGAAACCGAGAGATCTCTAAGGAGGCCCGTGCA	2750
Qy	1617	CGGCGTGTACTAGACCCGACGAGGACCTGTGTGGCGGAGATCCAGAAGCAGGCGCCACGA	1676
Db	2751	CGGCGTGTACTATGACCCCTCAAGGACCTGTATCGCCGAGATCCAGAAAGCAGGGGCAAGG	2810
Qy	1677	CCAGTGACCTTACAGATCTACCAAGGAGCCCTTCAAGAACCTGAAAGCCCGCAAGTACGC	1736
Db	2811	CCAGTGACCTTACAGATTTACCAAGGAGCCCTTCAAGAACCTGAAAGCCCGCAAGTACGC	2870
Qy	1737	CAGATGCGCACCGCCCAACCAACGAGCTGGAAGAGAGCTGACCGAGGCGCTGCAGAAAGAT	1796
Db	2871	CCGGAATGAGGGGTGCCACACTAACACGCTCAAGCAGCTGACCGAGGCGCTGCAGAAAGAT	2930
Qy	1797	CGCCATGGAAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCCATCCAGAA	1856
Db	2931	CACCAACGAAGCATCTGTGATCTGGGGAAGAATCTCTAAGTTCAAGCTGCCCATCCAGAA	2990
Qy	1857	GGAGACCTGGGAGACCTGTGTGACCAACTACTGGCAGCGCACCTGTGATCCCGAGTGGGA	1916

[illegible]

```

RESULT 11
US-09-936-572-12
; Sequence 12, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GS00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
; OTHER INFORMATION: optimised HIV-1 gagpo
US-09-936-572-12

```

	Query Match	66.9%	Score 1651.4;	DB 4;	Length 4642;
	Best Local Similarity	82.2%;	Pred. No. 1.2e-273;		
	Matches 1951;	Conservative 0;	Mismatches 406;	Indels 16;	Gaps 4;
Qy	12	CATGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAA	CATCTGTATGATGAGCGCGACAA	68	
Db	1421	CCTGGCTGAGGCCATGAGCCAGGTGA	CCAACTCCGTACCATCATGATGAGCGCGCAA	1480	
Qy	69	CTTCAAGGGGCCCAAGCGCATCATCAAGTGCCTTCAACTGCGGGAAGAGGGGCCACATCGC	128		
Db	1481	CTTTTCGAAACCAACGCAAGATCGTCAAGTGCCTTCAACTGTGGCAAGAAGGGCACACAGC	1540		

Qy	129	CCGCAACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGCGGCAAGGAGGCCACCA	188	Qy	1197	CCCCGAGATCGTGATCTACAGTACATGGAGCAGCTGTACGTGGGACGACCTGGAGAT	1256
Db	1541	CCGCAACTGCGAGGCGCCCTAGGAAAGAGGGCTGTGGAAGTGTGGAAGGAGGACCA	1600	Db	2620	CCCCGACATCGTATCTATCAGTACATGGATGAATGTGTACGTGGGCTCTGATAGAGT	2679
Qy	189	GATGAAGGACTGCACCGAGCCGAGGCCCAACTCTTCGCGAGGACCTGGCCCTTCCCCA	248	Qy	1257	CGGCCAGCACCGCCCAAGATTCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCAC	1316
Db	1601	AATGAAGATTTGTAATCAGAGACAGGCTAA-TTTTATTAGGAAGATCTGGCCCTTCCACA	1659	Db	2680	AGGCGAGACCGCAACAGATTCGAGGAGCTGCGCAGCACCTGTTGAGTGGGACTGAC	2739
Qy	249	GGCAAGGCCCGCGAGTTCCCGAGCGAGCAAGCCGCGCCAAAGCCGCCACAGCCGCGA	308	Qy	1317	CACCCCGCAGAAGAGCACCAAGAGGCCCTTCTGTGGATGGGCTACGAGCTGCA	1376
Db	1660	AGGCAAGGCCAGGGAAATTTCTTCAGAGCAGACAGAGCCAAAGCCGCCACCAAGAGA	1719	Db	2740	CACACCCGACAAGAAGCACCAAGAGGAGCTCTCTCTCTGGATGGGTTACGAGCTGCA	2799
Qy	309	GCTCAGGTGCGCG-----CGACAACCCCGCAGAGAGGCGCGCGCGAGCCGAGGG	362	Qy	1377	CCCCGACAAGTGGACCTGCGAGCCCATCGAGCTGCCCGAGAGGAGAGCTGGACCGTAA	1436
Db	1720	GCTTCAGGTTTGGGGAAGAGACAACATCTCTCTCAGAGCAGAGCCAAAGCCCGATAGACAAG	1779	Db	2800	CCCTGACAAATGGACCGGTGCGAGCTATCGTGTCTCCAGAGAAAGACAGCTGGACTGTCAA	2859
Qy	363	CA-----CCCTGAATTTCCCGAGATCACCTGTGCGAGCGCCCTCTGTGTGACATCAA	416	Qy	1437	CGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCCGGATCAA	1496
Db	1780	AATGTATCTTTAGCTTCCCTCAGATCACTCTTTTGGAGCGACCCCTCTGTCACAATAA	1839	Db	2860	CGACATACAGAGCTGTGTGGGAAAGTTGAACTGGGCCAGTCAAGATTTACCCAGGGATTAA	2919
Qy	417	GGTGGGCGGCAGATCAAGAGAGGCCCTGTGGAACACCGGCGCCGACACACCTGTCTGGA	476	Qy	1497	GGTGGCGCAGCTGTGCAAGCTGTGCGGCGCCAAAGCCCTGACCGACATCTGTGCCCT	1556
Db	1840	GATAGGGGGGAGCTCAAGAGGCTCTCTTGGAACACCGGAGCAGACCGTGTCTGGA	1899	Db	2920	GGTGAGGCGAGTGTGCAAACTCTCTCGCGGAACCAAGSCACTCACAGAGGTGATCCCCCT	2979
Qy	477	GGAGATGAGCTGCGCGGCAAGTGGAAAGCCCAAGATGATGCGCGCATGCGCGCTTCAT	536	Qy	1557	GACCGAGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGCGAGCCGCTGCA	1616
Db	1900	GGAGATGCTGTGCGAGCGCGTGGAAAGCCCAAGATGATCGGGGGAATCGCGCGTTTCAT	1959	Db	2980	AACCGAGAGGCCGAGCTCGAACTGGCAGAAAACCGAGAGATCTTAAAGGAGCCGCTGCA	3039
Qy	537	CNAGTGGCGCAGTACACACAGATCTGTATCGAGATCTGGGCAAGAGGCCATCGGCAC	596	Qy	1617	CGGCTGTACTACGACCCAGCAAGAGCTGTGGCGCGAGATCCAGAGCAGGSCCACA	1676
Db	1960	CNAGTGGCGCAGTATGACCAAGTCTCTCATCGAAATCTGGGCGCCACCAAGGCTATCGGTAC	2019	Db	3040	CGGCTGTACTATGACCCCTCCAAAGGACCTGATCGCCGAGATCCAGAGCAGGSCCAAGG	3099
Qy	597	CGTCTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGG	656	Qy	1677	CCAGTGGACCTTACAGATCTACAGAGAGCCCTTCAAGAACCTGAGAACCCGCAAGTACGC	1736
Db	2020	CGTCTGTGGGCCCCACACCCGTCAACATCATCGGAGGACACCTGTGTGACGAGATCGG	2079	Db	3100	CCAGTGGACCTTACAGATTTTACAGAGAGCCCTTCAAGAACCTGAGAACCCGSCNAGTACGC	3159
Qy	657	CTGACCCCTGAATTTCCCATACGCCCATCGAGACCGTGCCTGTGAAGCTGGAAGCCCGG	716	Qy	1737	CAAGATCGCACCGCCCAACCAACGACGTGAAGAGCTGACCGAGGCGGTGCAAGAT	1796
Db	2080	TTGCACGCTGAATTTCCCATTTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG	2139	Db	3160	CCGATGAGGGGTGCCACACACTAACGACGTCAAGCAGCTGACCGAGGCGGTGCAAGAGAT	3219
Qy	717	CATGGAAGCCCGCAGTGAAGAGTGGCCCTTACCGAGGAGAGATCAAGGCCCTGAC	776	Qy	1797	CGCATCGAGAGGATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAA	1856
Db	2140	GATGACAGCCCGAGGTCAAGCAATGGCCATTTCAGAGAGAGAGATCAAGGCATCGT	2199	Db	3220	CACCACCAAGAGCATCTGTGATCTGGGGAAAGACTCTTAAGTTCAAGCTGCCCATCCAGAA	3279
Qy	777	CGCCATCTGCGAGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGAGAACCC	836	Qy	1857	GGAGACCTGGGAGACCTGTGTGGACCCGACTACTGGCAGGCCACCTTGGATTCGCCGAGTGGGA	1916
Db	2200	GGAGTTTGACAGAGATGGAAGAGAGGGAATCTCCAGATTTGGGCTTGAAACCC	2259	Db	3280	GGAACTCTGGAAACCTGTGTGGACAGAGATTTGGCAGGCCACCTTGGATTCCTGAGTGGGA	3339
Qy	837	CTACAAACCCCGTGTTCGCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGT	896	Qy	1917	GTTCGTGAACACCCCGCCCTCTGTGTGAAGCTGTGTGTAACAGCTGAGAGAGGCCCATCAT	1976
Db	2260	GTAACAACGCGGTGTTCGCAATCAAGAAGAGGACTCGACAGAAATGGCGCAAGCTGGT	2319	Db	3340	GTTCGTCAACACCCCTCCCTCTGTGTGAAGCTGTGTGTACAGCTGGAGAGGAGGCCATAGT	3399
Qy	897	GGACTTCGGAGTGAACAGCGCACCCAGGACTCTGGAGGTGAGCTGGGCATCCC	956	Qy	1977	CGGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGGAGACCAAGATTCGCGCAAGGC	2036
Db	2320	GGACTTCGGAGTGAACAGCGCACCGCAAGACTTCTGGAGGTTTCAGCTGGGCATCCC	2379	Db	3400	GGGCGCGCAAACTTCTACGTGTGATGGGCGCGCTTAAAGGGAGACTTAAGCTGGGCAAGC	3459
Qy	957	CCACCCCGCGCTGAAGAAGAGAGAGCGTGCACCGTGTGACGCTGGGCGCAGCCTA	1016	Qy	2037	CGGCTACTGACCGACCGCGCGCGCAAGAGATCGTGAGCCTGTACCGAGACCAACCA	2096
Db	2380	GCACCCCGCAGGCTGAAGAAGAGAAATTCGTGACCGTACTGTGATGGGTGATGCCCTA	2439	Db	3460	CGGATAGTCACTAAACCGGGGACAGACAGAGGTGTGTACCTCTACTGACACCAACCA	3519
Qy	1017	CTTCAGCGTGCCTGAGCAGGACTTCCGCAAGTACACCGCCTTACCAATTCGCCAGCAT	1076	Qy	2097	GAAGACCGAGCTGAGGCCATCCAGTGGCCCTGTGACGAGACGCGGACGAGGTGAACAT	2156
Db	2440	CTTCTCTGTTCCCTGCGACGAGACTTCAGGAAGTACACTGCTTCAATTCCTTCGAT	2499	Db	3520	GAAGACTGAGCTGAGGCCATTTACCTCGCTTTTGAGGACTCTCGGCGCTGGAGGTGAACAT	3579
Qy	1077	CAACAACGAGACCCCGGATTCGCTACAGTACAGCTGTGCGCCCGAGGGCTGGAAGGG	1136	Qy	2157	GTGACCGACAGCCAGTACCGCCCTGGGCATCATCCAGCGCCCGCCGACAGAGCGAGAG	2216
Db	2500	CAACAACGAGACACCGGGGATTCGATATCAGTACAGCTGTGCGCCCGAGGGCTGGAAGG	2559	Db	3580	CGTGACAGACTCTCAGTATGCCCTGGGCATCATTTCAAGCCCGACGACAGAGTGTGATC	3639
Qy	1137	CAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGCGCTTCCGCGCCCGCAA	1196	Qy	2217	CGAGCTGGTGAAACAGATCATCGAGCAGCTGTATCAAGAGGAGAGGTGTACCTGTGAGCTG	2276
Db	2560	CTCTCCCGCAATCTTCAGAGTAGCATGACCAAAATCTGTGAGCTTTCGCGCAAAACAGAA	2619	Db	3640	CGAGCTGGTCAATCAGATCATCGAGCAGCTGTATCAAGAGGAAAGGTCTATCTGGCGCTG	3699
				Qy	2277	GGTGGCCCGCCCAAGAGGGCATCGCGCGCAACGAGCAGATCGACAAAGCTGTGTGAGCAAGGG	2336



Db 3700 GGTACCGCGCCCAACAAAGGCAATGGCGCAATGAGCAGTGCACAAGCTGCTCGGCTGG 3759  
 QY 2337 CATCCGCAAGGTGCTGTTCTGACGCGCATCGA 2369  
 Db 3760 CATCAGAAGGTGCTATTCTCGATGGCATCGA 3792

RESULT 12

US-09-552-950-5  
 ; Sequence 5, Application US/09552950  
 ; Patent No. 6541248  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Oxford Biomedica (UK) Limited  
 ; TITLE OF INVENTION: Anti-Viral Vectors  
 ; FILE REFERENCE: 674524-2004  
 ; CURRENT APPLICATION NUMBER: US/09/552,950  
 ; CURRENT FILING DATE: 2000-04-20  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 9772  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:psYNGP  
 US-09-552-950-5

Query Match 66.9%; Score 1651.4; DB 4; Length 9772;  
 Best Local Similarity 82.2%; Pred. No. 1.4e-273;  
 Matches 1951; Conservative 0; Mismatches 406; Indels 16; Gaps 4;

QY 12 CATGGCGGAGCCATGAGCAGG---CCACGAGGCCAATCCTGATGAGCGGAGCA 68  
 Db 2193 CTTGAGGCGCCATGAGCAGGTGACCAATCCGCTACCATCATGATGAGCGGCGCA 2252  
 QY 69 CTTCAAGGGCCCCAAGCGGCATCAAGTGTCTCAACTGCGGCAAGGAGGCCACATGCG 128  
 Db 2253 CTTTCGACCAACGCAAGATCTCAAGTGTCTCAACTGTTGGCAAGAGGGCACAGC 2312  
 QY 129 CCGCAACTGCGGCGCCCCCGCAAGAGGGCTGCTGGAAAGTGGCGCAAGGAGGCCACCA 188  
 Db 2313 CCGCAACTGCGGCGCCCCTAGGAAAAGGGCTGTGGAAATGTGAAAAGGAGGACACCA 2372  
 QY 189 GATGAGGACTGACCGAGCGCCAGGCAACTTCTTCGCGAGGACTGCGCTTCCCCCA 248  
 Db 2373 AATGAAAGATTGTACTGAGAGAGAGGCTAA-TTTTTTAGGGAAGATCTGGCCCTTCCACA 2431  
 QY 249 GGGCAAGGCGCGGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGACCCCAACGAGCGCGA 308  
 Db 2432 AGGGAAGGCCAGGGAATTTTCTCAGAGCAGACCGAGGCCAAGCCCCACAGAGAGA 2491  
 QY 309 GCTGCAAGGTGCGCG-----CGACAAACCCCGAGCGAGGCGCGCGCGAGCGCAGGG 362  
 Db 2492 GCTTCAGGTTTGGGAAGAGACAACTCCCTCTCAGAGCAGGAGCGATAGACAAGG 2551  
 QY 363 CA-----CCCTGAACTTCCCGCAGATCACCTGTGCGAGCGCCCTTGTGATGATCAA 416  
 Db 2552 AACTGTATCTTTAGTCTTCCCTCAGATCACTTTTGGCAGCAGCCCTCGTCAATAAA 2611  
 QY 417 GGTGGCGGCGAGATCAAGAGAGCCCTGTGGACACCGCGCGCGAGCAGACCCGTGTGGA 476  
 Db 2612 GATAGGGGGCAGCTCAAGAGAGCTCTCTTGGACAACCGAGCAGACGACACCGTGTGGA 2671  
 QY 477 GGAGATGAGCTTCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGTTTAT 536  
 Db 2672 GGAGATGTCTTCCGAGCGCTGGAAGCCGAAGATGATCGGCGGAATCGGCGGTTTAT 2731  
 QY 537 CAAGTGGCGGAGTACGACAGATCTGTGATCGAGATCTGGCGCAAGAGGCCATCGGCAC 596  
 Db 2732 CAAGGTGGCGGAGTATGACAGATCTCTCATCGAAATCTGGCGCAACAGGCTATCGGTAC 2791  
 QY 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGTGACCCAGCTGGG 656

Db 2792 CGTGTGTGGGCGCCACACCCGTCACATCATCGGACGCAACTGTGTGACGAGATCGG 2851  
 QY 657 CTGCACCTCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGAACTGAGCTGAGCCGG 716  
 Db 2852 TTGCACGCTGAACTTCCCATTTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCG 2911  
 QY 717 CATGGACGCGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAATCAAGGCCCTGAC 776  
 Db 2912 GATGGACGCGCCGAAAGGTCAAGCAATGGCCATTGACAGAGGAGAAATCAAGGCATGGT 2971  
 QY 777 CGCCATCTCGAGGAGATGGAGGCAAGGCAAGATCAACAAATCGGCCCCGAGAAACCC 836  
 Db 2972 GGAGATTTGCACAGAGATGGAAAAGGAAATCTCCAAGATGGGCTCGAGAACCC 3031  
 QY 837 CTACAAACACCCCGTCTTCCCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGGT 896  
 Db 3032 GTACAAACACGCGGTGTTGCAATCAAGAAAGAGGACTCTGACGAAATGGCGCAAGCTGGT 3091  
 QY 897 GGACTTCCGCGAGCTGAAACAAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCATCCC 956  
 Db 3092 GGACTTCCGCGAGCTGAAACAAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCATCCC 3151  
 QY 957 CCACCCCGCGGCTCTGAAAGAAAGAGCGTGAACCGTGTGAGCGTGGCGGCAAGCCATA 1016  
 Db 3152 GCACCCCGCGAGGCTGAAAGAAAGAAATCCGTTGACCGTACTGGATGTGGGTGATGCTA 3211  
 QY 1017 CTTGAGGTGCGGCTTGGACGAGCACTTCGCAAGTACACCGCTTCAACCATCCCAGCAT 1076  
 Db 3212 CTTTCTCGTTCGCTGGACGAGACTTTCAGGAAAGTACACTGCTTTCACAAATCCCTTCGAT 3271  
 QY 1077 CAAACAGAGACCCCGGATCGCTACCAAGTAAACAGTCTGCTGCCCGAGGGCTGGAAGGG 1136  
 Db 3272 CAAACAGAGACACCGGGATTCGATATCAGTACAGTGTCTGCCCGAGGGCTGGAAGGG 3331  
 QY 1137 CAGCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCAA 1196  
 Db 3332 CTCTCCGCAATCTTCAGAGTAGCATGACCAAAATCTCGAGGCTTTCGCGCAACAGAA 3391  
 QY 1197 CCCGAGATCGTGTCTACAGTACATGACGACCTGTAGTGGGAGCGACCTGGAGAT 1256  
 Db 3392 CCCGAGATCGTGTCTATCAGTACATGATGATCTTGTACGTGGGCTCTGATCTAGAGAT 3451  
 QY 1257 CGCCAGCAGCCGCGCAAGATGAGGAGCTGCGCAAGACCTGTGCTGGGCTTTCAC 1316  
 Db 3452 AGGGAGACCCGACCAAGATCGAGGAGCTGCGCAGACCTGTTGAGGTGGGACTGAC 3511  
 QY 1317 CACCCCGCAACAGAGCAACAGAGAGCGCCCTTCTGTGTGATGGGCTACGAGCTGCA 1376  
 Db 3512 CACACCCGCAACAGAGCAACAGAGAGCGCTTCTCTGTGATGGGTTACGAGCTGCA 3571  
 QY 1377 CCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGGAGCTGGACCGTGA 1436  
 Db 3572 CCCTGCAAAATGGACCGTGCAGCTTATCGTGTGCGCAGAGAAAGACAGCTGGACTGCA 3631  
 QY 1437 CGACATCCAGAGCTGTGGCAAGCTGAACCTGGCGCAGCAGATCTATCCCGCGCATCAA 1496  
 Db 3632 CGACATACAGAGCTGTGGGAGTTGAACTGGGCGAGTCAGATTTACCCAGGGATTAA 3691  
 QY 1497 GGTGCGCCAGCTGTGAAGCTGTGCGCGCGCAAGGGCCCTCAACGACATCGTGGCCCT 1556  
 Db 3692 GGTGAGGAGCTGTGCAAACTCTCCCGCGAAACCAAGGCCACTCACAGAGGTGATCCCCCT 3751  
 QY 1557 GACCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGGAGATCTTGGCGAGCCCGTGA 1616  
 Db 3752 AACCGAGGAGCGGAGCTCGAACTGGCAGAAAACCGAGAGATCTTAAAGAGCCCGTGA 3811  
 QY 1617 CGCGCTGTACTAGACCCAGCAAGCACTGTGTGGCGAGATCCAGAAAGCAGGCGCACGA 1676  
 Db 3812 CGCGCTGTACTATGAACCCCTCCAAAGNACTGTATCGCGAGATCCAGAGCAGGGCAAG 3871  
 QY 1677 CCAGTGGACCTACAGATCTTACAGAGCCCTTCAAGAACTCTGAAGACCGGCGAGTACGC 1736

Db 3872 CCAGTGGACCTTACAGATTACAGAGGCCCTTCAAGAACTGGAAGCCGCAAGTAGGC 3931  
Qy 1737 CAAGATGGCACCGCCACACACGACGTAAGAGCTGACCGAGCGCTGCGAGAGAT 1796  
Db 3932 CCGATGAGGGGTGCCACACATCAACGACGCTCAAGCAGCTGACCGAGCGCTGCGAGAGAT 3991  
Qy 1797 CGCCATGGAGGACATCGTATCTGGGGCAGAGACCCCAAGTTCCGCTGCGCATCCAGAA 1856  
Db 3992 CACCACCGAAAGCATCGTATCTGGGGAAGACCTCTAAGTTTCAAGCTGCGCCATCCAGAA 4051  
Qy 1857 GGAGACCTGGGAGACCTGGTGGACCGGACTACTGGCAGGCCACCTGGATCCCGAGTGGGA 1916  
Db 4052 GGAACCTGGGAACCTGGTGGACAGATATGGCAGGCCACCTGGATCTCTGATGGGA 4111  
Qy 1917 GTTCGTGAACACCCCGCCCTCTGGTGAAGCTGTGTATCCAGCTGAGAGAGGCCATCAT 1976  
Db 4112 GTTCGTCAACACCCCTCCCTGGTGAAGCTGTGTATCCAGCTGAGAGAGGCCATCAT 4171  
Qy 1977 CGGGCCGAGACCTTCTACGTGGAGCGGGCGGCCCAACCGCGAGACCAAGATCGGCAAGGC 2036  
Db 4172 GGGCGCCGAAACCTTCTACGTGGATGGGGCGGCTTAAAGGGAGACTAAGCTGGGCAAGC 4231  
Qy 2037 CGGTACGTGACCGACCGGGCGCGCAGAGATCGTGAAGCTGACCGAGACCAACCA 2096  
Db 4232 CGGATACGTCACTAACCGGGCGCAGAGAGAGTTGTCACTCTGACACCAACCA 4291  
Qy 2097 GAAGACCGAGCTGACGACCATCCAGCTGGCCCTCCAGGACAGCGCGAGGTGAACAT 2156  
Db 4292 GAAGACTGAGCTGACGACCATTTACCTCGCTTTCGAGGACTCGGGCTGGAGGTGAACAT 4351  
Qy 2157 CGTGACCGACGACGCTGAGTATGCTTGGGCGCATCAATCAAGCGCCAGCCAGACGAGTGAGTC 4411  
Db 4352 CGTGACGAGCTCTCAGTATGCTTGGGCGCATCAATCAAGCGCCAGCCAGACGAGTGAGTC 4471  
Qy 2217 CGAGCTGCTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTG 2276  
Db 4412 CGAGCTGCTCAATCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTATCTGCGCTG 4471  
Qy 2277 GGTGCCCGCCACAAAGGGCATCGCGCAACGAGCAGATCGACAAAGCTGTGAGCAAGGG 2336  
Db 4472 GGTACCCGCGCCACAAAGGCTTGGCGCAATGAGCAGCTCGACAAAGCTGTGCTCGGCTGG 4531  
Qy 2337 CATCCGCAAGGTGCTGTTCTGGAGCGGCATCGA 2369  
Db 4532 CATCAGGAAGGTGCTATTTCTGGATGTCATCGA 4564

## RESULT 13

US-09-872-733A-6  
; Sequence 6, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; TITLE OF INVENTION: SIV ENV GENES  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872.733A  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 8366  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
; OTHER INFORMATION: of the construct pCMVgagpolenkan containing a CMV  
; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin  
; OTHER INFORMATION: resistance gene

## US-09-872-733A-6

Query Match 64.5%; Score 1592.8; DB 4; Length 8366;  
Best Local Similarity 80.7%; Pred. No. 1.4e-263;  
Matches 1914; Conservative 0; Mismatches 442; Indels 16; Gaps 4;  
Qy 14 TGGCGGAGGCGCATGAGCCAGCCACACAGC---GCCAATCTCTGATGCGAGCGCAGCAACT 70  
Db 1857 TGGCGGAGGCGCATGAGCCAGGTGACGAACCTCGGCGACCATTAATGATGCGAGAGGCAACT 1916  
Qy 71 TCAAGGGCCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGGAGGGCCACATCTCGCCC 130  
Db 1917 TCCGGAACCGAGCGGAAGATCGTCAAGTCTTCAATTTGGCAAGAGAGGGCACACCGGCA 1976  
Qy 131 GCAATCCCGCGCCCCCGGCAAGAGGGCTCTCGAAAGTGCAGGAGAGGGCCACACAGA 190  
Db 1977 GGAATCGCGGGCCCCCGGGAAGAGGGCTGTGTGAAATGTGGAAGAGGAAGACACAAA 2036  
Qy 191 TGAAGGACTGCACCGAGCGCGCCCAACTTCTTCCGCGAGGACTTGGCCTTCCCCCAGG 250  
Db 2037 TGAAGGATTTGACTGAGAGACAGGCTAA-TTTTTTTAGGGAAGATCTGGCCTTCTTACAG 2095  
Qy 251 GCAAGGCGCGGAGTTCCTCCAGCGAGCAGAAACCGGCCCAACAGCCCAACAGCGCGCAGC 310  
Db 2096 GGAAGGCGAGGGAATTTTCTTCAAGCAGACAGCCAGCCACAGCCCGCCACGAGAGAGC 2155  
Qy 311 TGCAGGTGCGCGG-----CGACAAACCCCGCAGAGCGCGCGCGAGCGCCAGGGCA 364  
Db 2156 TTCAGGTCTGGGGTAGAGACAACTCTCCCTCAGAAGCAGAGAGCGCGATAGACAAGAA 2215  
Qy 365 -----CCCTGNACTTCCCCCAGATCACCTGTGTGGAGCGCCCTTGTGAGCATCAAGG 418  
Db 2216 CTGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCAAGTAAGGA 2275  
Qy 419 TGGCGCGCAGATCAAGGAGCGCTGTGACACCGCGCGCAGCAGACACCGTGTGAGG 478  
Db 2276 TCGGGGGGCAACTCAAGAGAGCGCTGTCTGATACAGAGCAGATGATACAGTATTAGAG 2335  
Qy 479 AGATGAGCCTGCCCGGCAAGTGGAAAGCCCAAGATGATCGCGGATCGCGGCTTTCATCA 538  
Db 2336 AATGAGTTTCCAGGAAGATGGAAACCAAAATATGATAGGGGGATCGGGGCTTTCATCA 2395  
Qy 539 AGTGCGCGCAGTAGCAGCATCTGATCGAGATCTCGGCGCAAGAGCCATCGCACCG 598  
Db 2396 AGGTGAGGCGATACGACCATCTCTAGAAATCTGTGGAATATAAGTATAGGTACAG 2455  
Qy 599 TGCTGATCGCGCCACCGCTGAAACATCTCGGCGCAACATCTGACCCAGCTGGGCT 658  
Db 2456 TATTAGTAGGACCTACCTGTCAACATATTGGAAGAAATCTGTTGACCCAGATCGGCT 2515  
Qy 659 GCACCTCTGAATTTCCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCA 718  
Db 2516 GCACCTTGAATTTCCCATCAGCCCTATTGAGACGGTCCCGTGAAGTTGAAGCCGGGA 2575  
Qy 719 TGAACGCGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCG 778  
Db 2576 TGAACGCGCCCAAGGTCAAGCAATGGCCATTTGACGAAAGAGAGATCAAGGCCCTTAGTCG 2635  
Qy 779 CCATCTCGAGGAGATGGAAGAGGCGCAAGATCACCAAGATCGGCCCGCCAGAACCCCT 838  
Db 2636 AATCTGTACAGAGATGGAGAGAGAGAGATCAGCAAGATCGGCTGAGAACCCCT 2695  
Qy 839 ACAACACCCCGTGTTCGCCCATCAAGAAAGAGGACAGACCAAGTGGCGGCAAGCTGGTG 898  
Db 2696 ACAACACTCCAGTCTTCGCAATCAAGAAAGAGGACAGTACCAGTGGAGAAAGCTGGTG 2755  
Qy 899 ACTTCCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGACAGTGGGCAATCCCC 958  
Db 2756 ACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTTCTGGGAAGTTTCAAGTGGGCAATCCCC 2815  
Qy 959 ACCCGCGCGCTCAAGAAAGAGAGAGCGTGAACCGTGTGGAGCTGGCGGACCGCTACT 1018  
Db 2816 ATCCCGCTGGGTGGAAGAAAGAAAGTCAAGTGAACAGTGTGGATGTGGGTGATGCCCTACT 2875

```
QY 1019 TCAGCGTGCCTTGGAGGACTTCGCAAGTACACCGCTTCCATCCCCAGCATCA 1078
Db 2876 TCTCCGTTCCCTTGGACGAGGACTTCAGGAAGTACACTGCTTTCCAGATACCTAGCATCA 2935
QY 1079 ACAACAGAGACCCCGGCATCCGCTACCACTAGTACAACTGTGCTGCCCGCAGGGCTGGAAGGGCA 1138
Db 2936 ACACAGAGACACAGGCATCCGCTACAGTACAACTGTGCTGCCCGCAGGGCTGGAAGGGAT 2995
QY 1139 GCCCCAGCATCTTCCAGAGACAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCCGCAACC 1198
Db 2996 CACCAGCCATCTTTCAAAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCAAGCAAAACC 3055
QY 1199 CCGAGATCTGTATCTACAGTACATGAGCGACTGTAGCTGGCGAGGACCTTGGAGATCG 1258
Db 3056 CAGACATCTGTATCTATCATATAGTACAGCGACTCTACGTAGGAAGTGAACCTGGAGATCG 3115
QY 1259 GCCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGCTGCGCTGGGGCTTCAACCA 1318
Db 3116 GGCAGCACAGGACCAAGATCGAGAGCTGAGACAGCATCTGTTGAGTGGGACTGACCA 3175
QY 1319 CCCCCACAAGAACACCAAGAGGAGCCCCCTTCTGTGTGATGGGCTACGAGCTGAC 1378
Db 3176 CACCAGACAAGAACCAAGAGAACCTCCCTTCTGTGTGATGGGCTACGAACCTGCATC 3235
QY 1379 CCGAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGGAGAGAGCTGGACCGTGAACG 1438
Db 3236 CTGACAAGTGGACAGTGCAGCCCATCGTGTGCTGAGAGGAGACGTGACTGTGAACG 3295
QY 1439 ACATCAGAAAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCGGGCATCAAG 1498
Db 3296 ACATCAGAAAGCTGTGGGCAAGTGAACCTGGGCCAGCCAGATCTACCCAGGCATCAAG 3355
QY 1499 TGCGCAGCTGTCAAAGCTGTGCGCGGCGCCAAAGCCCTTGACCGACATCTGCCCCCTGA 1558
Db 3356 TTAGGCAGCTGTCAAAGCTGTTCGAGGAACCAAGGCACCTGACAGAAGTGAATCCCACTGA 3415
QY 1559 CCGAGAGCCCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGGAGCCGCTGCAG 1618
Db 3416 CAGAGGAAGCGAGCTAGAACTGGCAGAGAACCGAGAGATCTCTGAAGGAGCCAGTACATG 3475
QY 1619 GCCTGTACTACGACCCCGACGAGACTGTGTGCGCGAGATCCAGAAGCAGGGCCAGACC 1678
Db 3476 GAGTGTACTACGACCCCGACGAGACTGTATCCAGAGATCCAGAGCAGGGCCAGGGCC 3535
QY 1679 AGTGGACCTACAGATCTTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738
Db 3536 AATGGACCTACCAAACTTACAGAGGCCCCCTTCAAGAACCTGAAGACAGGCAAGTACGCCA 3595
QY 1739 AGATGCGACCGGCCACACCAAGAGCTGAGAGAGCTGACCGAGGCGGTGCGAGAAGATCG 1798
Db 3596 GGATGAGGGGTGCCCCACACCAAGATGTGAAGCAGCTGACAGAGCAGTGCAGAAATCA 3655
QY 1799 CCATGAGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTTCCGCTGCCCATCCAGAAGG 1858
Db 3656 CCACAGAGAGCATCTGTATCTGGGGCAAGACTTCCCAAGTTTCAAGCTGCCCATACAGAGG 3715
QY 1859 AGACTTGGAGACCTGTGTGACCGCATCTA CTGGCAGGCGACCTGGATCCCGAGTGGGAGT 1918
Db 3716 AGACATGGGAGACATGTTGACCGAGTACTTGGCAAGCCACTTGGATCCCTGAGTGGGAGT 3775
QY 1919 TCGTGAACACCCCTTGTGATGAGCTGTGTGTACAGCTGGAGAGAGGAGCCCATCATCG 1978
Db 3776 TCGTGAACACCCCTTGTGTAAGTATGTGGTATCAGCTGGAGAGAGAACCCCATCTGTGG 3835
QY 1979 GCGCCGAGACCTTCTACGTGGACGGCGCCCAACCGCAGACCAAGATCGGAAGGCCG 2038
Db 3836 GAGCAGAGACCTTCTACGTGTGATGGGCGACCAACAGGGAGACCAAGCTGGGCAAGGCAG 3895
QY 2039 GCTACGTGACCGACCGGGCGCGCAGAAAGATCTGTGAGCTGACCGCAGACCAACCAACGA 2098
Db 3896 GCTACGTGACCAACCGGAGGACGACAGAAAGTGTGTGACCTCTGACTGACACCAACCAACGA 3955
```

```
QY 2099 AGACGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGCAGAGGTGAACATCG 2158
Db 3956 AGACTGAGCTGCAGGCATCTACCTAGCTCTGCAGACAGCGGACTGGAAGTGAACATCG 4015
QY 2159 TGACCGACAGCCAGTAGTACGCCCTTGGGCATCATCCAGGCCCGCAGCCGACAGAGGAGCG 2218
Db 4016 TGACGACTCACAGTAGTACGCACTGGGCATCATCCAGGACAAACAGACCAATCCGAGTCAG 4075
QY 2219 AGCTGTGAACAGCATCTGAGCAGCATCAAGAAGGAGAGGTGTACTGTAGCTGGG 2278
Db 4076 AGCTGTGAACAGCATCTGAGCAGCATCAAGAAGGAGAGGTGTACTGTGGCATGGG 4135
QY 2279 TGCCCCCCACAGAGGCGCATCGCGGCAACGAGCAGATCGACAAGCTGGTCAGCAAGGGCA 2338
Db 4136 TACCAGACACAAAGAAATTGGAGGAATGAACAGTAGATANAATTAGTCAGTGTCTGGGA 4195
QY 2339 TCCGCAAGGTGCTGTTCTTCCCTGGACGGCATCGAT 2370
Db 4196 TCCGGAAGGTGCTGTTCTTCCCTGGACGGATCGAT 4227

RESULT 14
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

Query Match 63.1%; Score 1557.4; DB 4; Length 4338;
Best Local Similarity 80.5%; Pred. No. 1.5e-257;
Matches 1914; Conservative 0; Mismatches 441; Indels 22; Gaps 7;

QY 14 TGCCGAGGCGCATGAGCCAGGCCACCAGC---GCCAACATCTGTATGTCAGCGCAGCAACT 70
Db 1085 TGCCGAGGCGCATGAGCCAGGTCAGCAACTCGCGCGACCAATAATGATGTCAGAGAGGCAACT 1144
QY 71 TCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTTCGGCAAGGAGGGGCCAATCGCCCC 130
Db 1145 TCCGGAACACAGCGGAAGATCGTCAAGTGTCTTAATTGTGGCAAGAGAGGGCAACCCGCA 1204
QY 131 GCAACTGCCGCGCCCCCGCAAGAGGGCTGTGGAAAGTTCGCGCAAGAGGGGCCACCAAGA 190
Db 1205 GGAACCTGCCGGGGCCCCCGGAAGAGGGCTGTTGGAAATGTGGAAAGGAGGACACCAAA 1264
QY 191 TGAAGATGTCACCGAGCGCCAGGCCAACACTTCTTCGCGAGGACCTGGCCCTTCCCCCAGG 250
Db 1265 TGAAGATTTGTACTGAGACAGAGCTTAA-TTTTTTAGGGGAAGATCTGGCCCTTCTCAAG 1323
QY 251 GCAAGGCCCGCGATTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCCCACCAGCGCGAGC 310
Db 1324 GGAAGGCCAGGGAATTTTCTTCAAGAGCAGACAGAGCCACAGCCCCCACCAGAGAGAGC 1383
QY 311 TGCAGGTGCGCGG-----CGAACACCCCGCAGCGAGGCGCGCGCGCAGCGCAGGGCA 364
```

Db 1384 TTCAGGTCGGGTAGAGACAACTCCCTCCCTCAGAAAGCAGGAGCCGATAGACAAAGGAA 1443  
Qy 365 -----CCCTGAACCTTCCCCAGATCAACCTGTGGCAGCGCCCTGTGTGAGCATCAAGG 418  
Db 1444 CTGTATCTTTAACTTCCCTCAGATCACTTTGGCAACGACCCCTGTGTACAGTAAGGA 1503  
Qy 419 TGGGGGGCCAGATCAAGAGGCCCTGTGTGGACACCGGGCCGACGACACCGTGTGGAGG 478  
Db 1504 TCGGGGGGCAACTCAAGGAAGCGTGTGTGATACAGGAGCAGATGATACAGTATTAGAA 1563  
Qy 479 AGATGAGCTGCCCGGAAGTGAAGCCCAAGATGATCGCGGCAATCGCGGCTTCATCA 538  
Db 1564 AATGAGTTTCCAGGAAGATGGAACCAAAATGATAGGGGGATCGGGGGCTTCATCA 1623  
Qy 539 AGGTGCGCCAGTACGACAGATCTGTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACG 598  
Db 1624 AGGTGAGGACGTACGACAGATCTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 1683  
Qy 599 TGCTGATCGGCCCCACC-----CCGTGAACATCATCGGCCGCAACATGCTGACCCAGCTG 654  
Db 1684 TATTAGTAGGACCTACCTACACCTGTCAACATAATTGGAAGAAATCTGTGACCCAGATC 1743  
Qy 655 GCGTGACCTTGAATTCCTCCCATGAGCCCATCGAGCCGTCGCGTGAAGCTGAAGGCC 714  
Db 1744 GCGTGACCTTGAATTCCTCCCATGAGCCCTTGTGAGACGGTGCCTGTAAGTGAAGCCG 1803  
Qy 715 GCGTGAAGCGCCCAAGGTGAAGCAGTGGGCCCTTGACCGAGGAGAGATCAAGGCCCTG 774  
Db 1804 GGGATGAGCGGCCCAAGGTCAAGCAATGGCCATTGACAAAGAGAGATCAAGGCCCTTA 1863  
Qy 775 ACCGCGATCTCGGAGGAGTGAAGAGGAGGCAAGATCACAAAGATCGGCCCGAGAAC 834  
Db 1864 GTCGAAATCTGTACAGAGATGGAAGAGGAGGAGATCAGCAAGATCGGCCCTGAGAAC 1923  
Qy 835 CCCTACAAACCCCGTGTCCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTG 894  
Db 1924 CCCTACAAACTCTCAGTCTTCCCAATCAAGAGAGGAGCAGTACCAAGTGGAGAAAGCTG 1983  
Qy 895 GTGACATCTCCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954  
Db 1984 GTGACATCTCAGAGCTGAACAAAGAGACTCAGGACTTCTGGGAAGTTTACCTGGGCATC 2043  
Qy 955 CCCACCCCGCGCTGAAGAGAGAGAGCGTGAACCGTGTGGAGCTGGGCGAGGCC 1014  
Db 2044 CCACATCCCGTGGGTGAAGAGAGAGAGTCAAGTGAAGTGTGGTGTGGTGTGGTGTGGT 2103  
Qy 1015 TACTTCAGCTGCGCCCTGGAGAGGACTTCGCAAGTACACCGCTTCCACCATCCCGAGC 1074  
Db 2104 TACTTCTCCGTTTGGACGAGACTTCAGGAAGTACACTGCGCTTCCAGATACCTAGC 2163  
Qy 1075 ATCAACAGAGACCCCGGATCCGCTACAGGTACAAAGTGTGCGCCCGAGGCTGGAA 1134  
Db 2164 ATCAACAGAGACCCAGGATCCGCTACAGTACAAAGTGTGCGCCCGAGGATGAGAG 2223  
Qy 1135 GGCAGCCCGAGATCTTCAGAGAGAGATGACCAAGATCTGGAAGCCCTTCCGCGCCGCG 1194  
Db 2224 GGATCACAGCCATCTTCAAGAGAGATGACCAAGATCTCGAGGACCTTCCGCAAGCAA 2283  
Qy 1195 AACCCGAGATCTGTATCTACAGTACATGAGGACCTGTAGCTGGGCGAGCGCTGGAG 1254  
Db 2284 AACCCAGATCTGTATCTATCAGTACATGAGGACCTTACGTTAGGAAGTGTATCTGGAG 2343  
Qy 1255 ATC-GGGCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTGCTGCGCTGGGGCTT 1313  
Db 2344 ATCGGGGAGCAGCAGGACCAAGATCGAGGAGCTGAGACAGATCTGTGAGGTGGGACT 2403  
Qy 1314 CACCACCCCGAGCAAGAGACCAAGAGAGGCCCTTCTGTGTGATGGGCTACGAGCT 1373  
Db 2404 GACCACCAAGAGAGCAAGAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2463  
Qy 1374 GCACCCCGAGAGAGTGAAG 1433  
Db 2464 GCATCTGACAAAGTGAAG 2523

Qy 1434 GAACGACATCCAGAAAGCTGTGGCAAGCTGAACTGGGCCAGCCAGATCTTACCCCGGAT 1493  
Db 2524 GAACGACATCAGAAAGCTGTGGCAAGCTGAACTGGGCCAGCCAGATCTTACCCAGGAT 2583  
Qy 1494 CAAGGTGGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCCCTGAGCCGACATCGTGCC 1553  
Db 2584 CAAGGTGGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCCCTGAGCCGACATCGTGCC 2643  
Qy 1554 CCTCAGCAGAGAGCCGAGCTGGAGCTGGCCGAGAACCCGCGAGATCTCTGCGCGAGCCCGT 1613  
Db 2644 ACTCAGAGAGAGAGAGCTAGAACTGGCAGAGAACCCGAGAGATCTCTGAGAGAGCCAGT 2703  
Qy 1614 GCACGGGTGTACTACGACCCCAAGAGAACCTGGTGGCCGAGATCTCAGAAAGAGGGCCA 1673  
Db 2704 ACATGGAGTGTACTACGACCCCAAGAGAACCTGTATCGCAGAGATCCAGAAAGAGGGCA 2763  
Qy 1674 CGACAGTGGACCTTACAGATCTACAGGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTA 1733  
Db 2764 AGGCCAATGGACCTTACCAATCTTACAGGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTA 2823  
Qy 1734 CGCCAAGATGGCACCCGCCACACAGCTGAAGCAGCTGACCGAGGCCGCTGCAGAA 1793  
Db 2824 CGCAAGATGGGGTGGCCACACAGCTGTGAGCAGCTGACAGGCGAGTGCAGAA 2883  
Qy 1794 GATCCCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCA 1853  
Db 2884 GATCACACAGAGAGCATCTGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943  
Qy 1854 GAAGGACCTGGGAGACCTGTGGACCGACTACTGCGAGGCCACCTGGATCCCGGAGTG 1913  
Db 2944 GAAGGAGACATGGGAGACATGTGTGGACCGAGTACTGGCAAGCCACCTGGATCCCTGAGTG 3003  
Qy 1914 GGAGTTCGTGAACACCCCTTGTGTGAACTGTGTGTGAGTGTGAGTGTGAGAGAGGCCAT 1973  
Db 3004 GGAGTTCGTGAACACCCCTTGTGTGAACTGTGTGTGAGTGTGAGTGTGAGAGAGGCCAT 3063  
Qy 1974 CATCGGCCGAGACCTTCTACGTGTGAGCGGCCCGCCAAACCGCGAGACCAAGATCGGCAA 2033  
Db 3064 CGTGGGAGCAGAGACCTTCTACGTGTGATGGGGCAGCCAAACAGGGAGACCAAGCTGGGCAA 3123  
Qy 2034 GCGCGGTACTGTGACACCGGGGCCGCGAGAGATCTGTGAGCTGTGAGCCGAGACCAACCA 2093  
Db 3124 GCGAGGTACTGTGACCAACCGAGGAGCAGCAAGAAAGTGTGTGACCTGACTGACACCAACCA 3183  
Qy 2094 CCAGAGACCGAGCTGCAAGGCTTCCAGCTGGCCCTTCAGGACAGCGGACGCGAGGTGAA 2153  
Db 3184 CCAGAGACTGTGAGTGTGAGGCTTACCTAGCTTCTGCAAGACAGCGGACTGTGAGAGTGA 3243  
Qy 2154 CATCGTACCGACAGCCAGTACGCTTGGGCTATCATCCAGGCCCGAGCCCGACAGAGCGA 2213  
Db 3244 CATCGTACAGACTTCAAGTAGG-CATGGGCTATCATCCAAAGCAACACAGACCAATCCGA 3302  
Qy 2214 GAGGAGCTGTGAGACAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACTTGAG 2273  
Db 3303 GTGAGAGCTGTGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACTTGCC 3362  
Qy 2274 CTGGGTCCCGCCCAAGGCGCATCGCGGCCCAAGCAGCAGATCGCAAGCTGGTGTGAGCAA 2333  
Db 3363 ATGGGTACAGACACAAAGAAATTTGAGGAAATGAAACAGTAGATTAATTTAGTCAGTGC 3422  
Qy 2334 GGGCATCCGCAAGGTGTCTTCTTGGACCGGCTCGAT 2370  
Db 3423 TGGGATCCGGAAGGTGTCTTCTTGGACCGGATCGAT 3459

RESULT 15  
US-09-952-060-1  
; Sequence 1, Application US/09952060  
; Patent No. 6733993  
; GENERAL INFORMATION:  
; APPLICANT: Emili, Emilio A.  
; APPLICANT: Youil, Rima

; APPLICANT: Bett, Andrew J.  
; APPLICANT: Chen, Ling  
; APPLICANT: Kaslow, David C.  
; APPLICANT: Shiver, John W.  
; APPLICANT: Toner, Timothy J.  
; APPLICANT: Casimiro, Danilo R.  
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  
; TITLE OF INVENTION: MODIFICATIONS  
; FILE REFERENCE: 20747Y  
; CURRENT APPLICATION NUMBER: US/09/952,060  
; CURRENT FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: PCT/US01/28861  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/317,814  
; PRIOR FILING DATE: 2001-09-07  
; PRIOR APPLICATION NUMBER: 60/279,056  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/233,180  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2577  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Codon optimized DNA encoding modified wt pol  
US-09-952-060-1

Query Match 54.1%; Score 1336.4; DB 4; Length 2577;  
Best Local Similarity 86.7%; Pred. No. 8.3e-220;  
Matches 1472; Conservative 0; Mismatches 226; Indels 0; Gaps 0;  
  
QY 672 CCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCGCGCATGAGCGGCCCAA 731  
DB 15 CCCCATCTCCCCATTGAGACTGTGCCTGTGAAGCTGAAGCTGGCATGGATGGGCCCAA 74  
  
QY 732 GGTGAAGCAGTGGCCCTGACCGAGGAGATCAAGGCCCTCACGCCCATCTGCGAGGA 791  
DB 75 GGTGAAGCAGTGGCCCTGACTGAGGAGAGATCAAGGCCCTGGTGAATCTGTCACTGA 134  
  
QY 792 GATGAGGAAGAGGCGCAAGATCAACAGATCGGCCCGCGAACCCTTACAACACCCCGT 851  
DB 135 GATGAGGAAGAGGCGCAATCTCCAAGATTGGCCCGCGAACCCTTACAACACCCCTGT 194  
  
QY 852 GTTCGCCATCAAGAAGAGACAGCACAAAGTGGCGCAAGCTGGTGAATCTCGCGAGCT 911  
DB 195 GTTTGCCATCAAGAAGAGACTCCACAAAGTGGGAAAGCTGGTGAATCTCGAGGAGCT 254  
  
QY 912 GAACAAGCGCACCCAGGACTTCTGGGAGGTGACGCTGGGATCCCCACCCCGCGGCCCT 971  
DB 255 GAACAAGAGGACCCAGGACTTCTGGGAGGTGACGCTGGGATCCCCACCCCGCTGCCCT 314  
  
QY 972 GAAGAAGAAAGAGCGTGACCGTGTCTGGAAGTGGGCGAGCCCTACTTTAGCGTGGCCCT 1031  
DB 315 GAAGAAGAAAGTGTCTGTGACTGTCTGTGATGTGGGAGTGCCTACTTCTGTGCCCT 374  
  
QY 1032 GGAAGAGAACTTCGCGAAGTACACCGCTTTCACATCCCCAGCATCAACAAAGAGACCCC 1091  
DB 375 GGATGAGGACTTCAGGAAGTACACTGTGCTTTCACCATCCCCCTCCATCAACAAATGAGACCCC 434  
  
QY 1092 CGGCATCCGTACAGTACAACTGTGCCCCAGGCTGGAAGGCGAGCCCGAGCATCTT 1151  
DB 435 TGGCATCAGGTAACAGTACAACTGTGCCCCAGGCTGGAAGGCGTCCCCCTGCCATCTT 494  
  
QY 1152 CCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTTCGCGCCCGCAACCCCGAGATCGTGAT 1211  
DB 495 CCAGTCTCTATGACCAAGATCTCTGGAGCCCTTTCAGGAAGCAGAACCTTGACATGTGAT 554  
  
QY 1212 CTACAGATACATGGAACGACTGTGAGGCGAGGAGCTGGAGATCGGCAGCACCGCGC 1271  
DB 555 CTACAGATACATGGAACCTGTATGTGGGCTCTGACCTGGAGATTGGGCGAGCACAGGAC 614

QY 1272 CAAGATCGAGGAGCTGCGCAAGCACCTGTGTGGCTGGGGCTTTCACCAACCCCGACAGAA 1331  
DB 615 CAAGATTGAGGAGCTGAGGCGAGCACCTGTGTAGGTGGGGCTGACCAACCCCTGACAAGAA 674  
  
QY 1332 GCAACAAGAGGAGCCCTTCTGTGTGATGGGTACGAGCTGCAACCCCGACAAGTGGAC 1391  
DB 675 GCACCAAGAGGAGCCCTTCTGTGTGATGGGTATGAGCTGCAACCCCGACAAGTGGAC 734  
  
QY 1392 CGTGAGCCCATCGAGCTGCGCGAAGAGAGAGCTGGAACCGTGAACGACATCCAGAGCT 1451  
DB 735 TGTGAGCCCATTTGTCTGTGTGAGAGGACTCTCTGAGCTGTGAATGACATCCAGAGCT 794  
  
QY 1452 GGTGGCAAGCTGAACTGGGCGCAGCGAGATCTACCCCGGCATCAAGGTGCGCAGCTGTG 1511  
DB 795 GGTGGCAAGCTGAACTGGGCGCTCCCAATCTACCTTGGCATCAAGGTGAGGAGCTGTG 854  
  
QY 1512 CAAGCTGTGCGCGCGCAAGCCCTGACCGACATCTGTGCGCCCTGACCGAGAGGCGGA 1571  
DB 855 CAAGCTGTGAGGGGCAACCAAGGCCCTGACTGAGGTGATCCCCCTGACTGAGGAGGCTGA 914  
  
QY 1572 GCTGAGCTGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGGCGTGTACTACGA 1631  
DB 915 GCTGAGCTGCGCTGAGAACAGGAGATCTTGAAGAGAGCCCTGTGATGGGGTGTACTATGA 974  
  
QY 1632 CCCAGCAAGGACCTGTGTGCGCGAGATCCAGAGAGGCGCCACGACCAAGTGGACCTACCA 1691  
DB 975 CCGCTCCAGGACCTGATTTGCTGAGATCCAGAGAGGCGCCAGGCGCAGTGGACCTACCA 1034  
  
QY 1692 GATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACCGCAAGTGGCGCACCGC 1751  
DB 1035 AATCTACAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCGAGGATGAGGGGGC 1094  
  
QY 1752 CCACACCAAGCTGAAGCAGCTGACCGAGCGGTGCGAGAGATCGCCATCGAGAGCAT 1811  
DB 1095 CCACACCAATGATGTGAAGCAGCTGAGGTGTGCGAAGATCACCACCTGAGTCCAT 1154  
  
QY 1812 CGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAGGAGACCTTGGGAGAC 1871  
DB 1155 TGTGATCTGGGCAAGACCCCAAGTTCAAGCTGCCCATCCAGAGGAGACCTTGGGAGAC 1214  
  
QY 1872 CTGGTGAACGACTACTGCGAGCCCACTTGGATCCCGAGTGGGAGTGTGTGAACACCCC 1931  
DB 1215 CTGGTGAACGACTACTGCGAGCCCACTTGGATCCCGAGTGGGAGTGTGTGAACACCCC 1274  
  
QY 1932 CCGCTGTGTGAAGCTGTGTGTAACAGCTGAGAGAGGACCATCATCGGCGCCGAGACCTT 1991  
DB 1275 CCGCTGTGTGAAGCTGTGTGTAACAGCTGAGAGAGGACCATTTGTGGGGCTGAGACCTT 1334  
  
QY 1992 CTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGGCTACGTGACCGA 2051  
DB 1335 CTATGTGATGGGCTGCCAAGAGGAGACCAAGCTGGGCAAGGCTGGCTATGTGACCAA 1394  
  
QY 2052 CGGGGCGCGGAGAAATCGTGAGCTGAGCGAGACCAACAGAGAGCGAGCGAGTGTGAACCA 2111  
DB 1395 CAGGGGCGGCGAGAGGTTGTGACCTGACTGACCAACCAACCAAGAGAGCTGAGTCCA 1454  
  
QY 2112 GGCATTCAGCTGGGCGCTCGAGACAGCGAGCTGGAACATCGTGACCGACAGCCA 2171  
DB 1455 GGCATTCATCTGGGCGCTCGAGACTCTGGCTTGGAGGTGAACATTTGTGACTGACTCCCA 1514  
  
QY 2172 GTACGCGCTGGGCTCATCTCAGGCCCGAGCCCGCAAGAGAGCGAGCGAGTGTGAACCA 2231  
DB 1515 GTATGCGCTGGGCTCATCTCAGGCCCGAGCTGATCATCTGAGTCTGAGCTGTGAACCA 1574  
  
QY 2232 GATCATCGAGCTGATCAAGAGGAGAGGTTGTAACCTGAGCTGGGTGCGGCCCGACAA 2291  
DB 1575 GATCATCGAGCTGATCAAGAGGAGAGGTTGTAACCTGAGCTGGGTGCGGCCCGACAA 1634  
  
QY 2292 GGCATTCGGGCGCAAGCAGCATCCAGAGCTGTGAGCAAGGCGATCCGCAAGGTGCT 2351  
DB 1635 GGCATTCGGGCGCAATGAGCAGGTGCAAGCTGGTGTCTGTGGCATTCAGGAAGGTGCT 1694

Qy 2352 GTTCCTGGACGGCATCGA 2369  
|||||||  
Db 1695 GTTCCTGGATGGCATTGA 1712

Search completed: June 2, 2005, 03:12:53  
Job time : 401.948 secs



**This Page Blank (uspto)**

Result No.	Score	Query Match	Length	DB	ID	Description	
1	2046	83.1	2306	4	US-09-475-515-82	Sequence 82, Appl	
2	2025.2	82.2	2312	4	US-09-475-515-84	Sequence 84, Appl	
3	2019.2	82.0	2300	4	US-09-475-515-83	Sequence 83, Appl	
4	1942.6	78.9	4319	4	US-09-475-515-6	Sequence 6, Appl	
5	1878.8	76.3	2305	4	US-09-475-515-80	Sequence 80, Appl	
6	1852	75.2	2239	4	US-09-475-515-81	Sequence 81, Appl	
7	1651.8	67.1	4307	4	US-09-552-950-2	Sequence 2, Appl	
8	1632.6	66.3	4307	4	US-09-936-572-2	Sequence 2, Appl	
9	1624.6	66.0	4327	4	US-09-936-572-14	Sequence 14, Appl	
10	1624.6	66.0	4353	4	US-09-936-572-13	Sequence 13, Appl	
11	1624.6	66.0	4642	4	US-09-936-572-12	Sequence 12, Appl	
12	1624.6	66.0	9772	4	US-09-552-950-5	Sequence 5, Appl	
13	1566	63.6	8366	4	US-09-872-733A-6	Sequence 6, Appl	
14	1530.6	62.1	4338	4	US-09-872-733A-1	Sequence 1, Appl	
15	1309.6	53.2	2577	4	US-09-952-060-1	Sequence 1, Appl	
16	1309.6	53.2	2850	4	US-09-952-060-5	Sequence 5, Appl	
17	1307	53.1	4053	4	US-09-952-060-34	Sequence 34, Appl	
18	1304.8	53.0	2577	4	US-09-952-060-3	Sequence 3, Appl	
19	1304.8	53.0	2850	4	US-09-952-060-7	Sequence 7, Appl	
20	1304.8	53.0	38519	4	US-09-952-060-28	Sequence 28, Appl	
21	1203.6	48.9	9010	4	US-09-184-418C-8	Sequence 8, Appl	
22	1175.8	47.7	9913	4	US-09-827-688-11	Sequence 11, Appl	
23	1172.6	47.6	8972	4	US-09-184-418C-9	Sequence 9, Appl	
24	1165.4	47.3	8959	4	US-09-184-418C-11	Sequence 11, Appl	
25	1163.8	47.3	2467	4	US-09-872-733A-3	Sequence 3, Appl	
26	1142	46.4	8892	4	US-09-184-418C-4	Sequence 4, Appl	
27	1106.8	44.9	2601	3	US-09-117-217-7	Sequence 7, Appl	

398 GCCCTGCTGAGCATCAAGTGGCGCGCAGATCAAGGAGGCCCTGCTGACACACCGCG 457  
|||||  
241 GCCCTCTGGTGAACATCAGGATCGCGCGCCAGCTCAAGGAGCGCTGCTGACACCGCGCG 300  
|||||  
458 CCGACGACACCGTCTGGAGAGATGAGCTGCCCGGCAAGTGGAAAGCCCAAGATGATCG 517  
|||||  
301 CCGACGACACCGTCTGGAGAGATGAACCTGCCCGGCAAGTGGAAAGCCCAAGATGATCG 360  
|||||  
518 GCGGCATCGCGGGCTTCATCAAGTGGCGCCAGTACGACAGATCTGATCGAGATCTGCG 577  
|||||  
361 GCGGGATCGGGGGCTTCATCAAGTGGCGGAGTACGACAGATCTCCCGTGGAGATCTGCG 420  
|||||  
578 GCAAGAGGCGCATCGGCACCGTCTGATCGGCGCCACCCCGCTGAAACATCATCGGCGCA 637  
|||||  
421 GCCAAGGCGCATCGGCACCGTCTGATCGGCGCCACCCCGCTGAAACATCATCGGCGCA 480  
|||||  
638 ACATGCTGACCCAGCTGGGCTGCAACCTGAACTTCCCATCAGCCCATCGAGACCGTGC 697  
|||||  
481 ACCTGCTGACCCAGATCGGCTGCACCTGAACTTCCCATCAGCCCATCGAGAGGTCG 540  
|||||  
698 CCGTGAAGCTGAAGCCCGGCATGACCGCCCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGG 757  
|||||  
541 CCGTGAAGCTGAAGCCCGGCATGACCGCCCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGG 600  
|||||  
758 AGAAGATCAAGGCCCTGACCGCCATCTCGAGGAGATGGAGAGGAGGCAAGATCACCA 817  
|||||  
601 AGAAGATCAAGGCCCTGCTGGAGATCTGCACCGGATGGAGAGGAGGCAAGATCAGCA 660  
|||||  
818 AGATCGGCCCGAGAACCCCTTACAACACCCCGGTGTTGCCCATCAAGAAAGAGACAGCA 877  
|||||  
661 AGATCGGCCCGAGAACCCCTTACAACACCCCGGTGTTGCCCATCAAGAAAGAGACAGCA 720  
|||||  
878 CCAAGTGGCGAGCTGGTGGACTTCCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 937  
|||||  
721 CCAAGTGGCGAGCTGGTGGACTTCCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 780  
|||||  
938 AGGTGACGTGGGCATCCCGACCCCGCGCTGAAAGAGAGAGAGCGTGACCGTGC 997  
|||||  
781 AGGTGACGTGGGCATCCCGACCCCGCGCTGAAAGAGAGAGAGCGTGACCGTGC 840  
|||||  
998 TGGAAGTGGCGAGCGCTACTTACGCTGCGCCCTGGACGAGGACTTCCCGAAGTACACCG 1057  
|||||  
841 TGGACGTGGCGAGCGCTACTTACGCTGCGCCCTGGACAGGACTTCCCGAAGTACACCG 900  
|||||  
1058 CTTTACCATCCCGAGATCAACAACGAGACCCCGCGATCCGCTACAGTACAAAGTGC 1117  
|||||  
901 CTTTACCATCCCGAGATCAACAACGAGACCCCGCGATCCGCTACAGTACAAAGTGC 960  
|||||  
1118 TGCCCGAGGGCTGGAAGGGCAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGG 1177  
|||||  
961 TGCCCGAGGGCTGGAAGGGCAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGG 1020  
|||||  
1178 AGCCCTTCCGCGCGCAACCCCGAGATCGTGATCTACAGGCGCCCTCTGACGTGGGCA 1237  
|||||  
1021 AGCCCTTCCGCGCGCAACCCCGAGATCGTGATCTACAGGCGCCCTCTGACGTGGGCA 1080  
|||||  
1238 GCGACCTGGAGATCGGCCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCATCTGCTGC 1297  
|||||  
1081 GCGACCTGGAGATCGGCCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCATCTGCTGC 1140  
|||||  
1298 GCTGGGGCTTCCACACCCCGCAAGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357  
|||||  
1141 GCTGGGGCTTCCACACCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
|||||  
1358 GCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAG 1417  
|||||  
1201 GCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAG 1260  
|||||  
1418 GCTGGAGCTGCAAGCATCCAGAGCTGGTGGGCAAGCTGGAAGTGGGCGGCGGCGGCGGCT 1477  
|||||  
1261 GCTGGAGCTGCAAGCATCCAGAGCTGGTGGGCAAGCTGGAAGTGGGCGGCGGCGGCGGCT 1320  
|||||  
1478 ACCCGGCGCATCAAGTGGCGGCGGCTGTGCAAGCTGTGCGGCGGCGGCGGCGGCGGCTGACCG 1537  
|||||

Db |||  
1321 ACCGCGGCATCAAGTGAAGCAGCTGTGCAAGCTGTGCGCGCACCAAGCGCTCGACCG 1380  
Qy |||  
1538 ACATCTGTCCTCTGACCGAGGAGCGAGCTGAGCTGGCCGAGAACCGCGAGATCTCTGC 1597  
Db |||  
1381 AGGTGATCTCCCTCTGACCGAGGAGCGAGCTGAGCTGGCCGAGAACCGCGAGATCTCTGA 1440  
Qy |||  
1598 GCGAGCCCGTGCACGCGGTGATCTACGACCCAGCAAGGACCTGCTGGCGCGAGATCCAGA 1657  
Db |||  
1441 AGAGCCCGTGCACGAGGTGATCTACGACCCAGCAAGGACCTGCTGGCGCGAGATCCAGA 1500  
Qy |||  
1658 AGCAGGCGCCACGACCAAGTGAACCTTACCAGATCTTACCAGGAGCCCTTCAAGAACCTGAAGA 1717  
Db |||  
1501 AGCAGGCGCCACGAGCGCAGTGGACCTACCAGATCTTACCAGGAGCCCTTCAAGAACCTGAAGA 1560  
Qy |||  
1718 CCGGCAAGTACGCCAAGATCGGCACCGCCACACCAACGACGCTGAAGCAGCTGACCGAGG 1777  
Db |||  
1561 CCGGCAAGTACGCCCGCATGCGCGGCGCCACCAACGACGCTGAAGCAGCTGACCGAGG 1620  
Qy |||  
1778 CCGTGCAGAGATCGGCATGAGAGATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGC 1837  
Db |||  
1621 CCGTGCAGAGATGAGCACCAGAGATCTGTGATCTGGGGCAAGATCCCAAGTTCGAGC 1680  
Qy |||  
1838 TGCCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGCGAGGCCCACTTGGGA 1897  
Db |||  
1681 TGCCCATCCAGAAAGGAGACCTGGGAGCGCTGGTGGATGGAGTACTTGGCAGGCCACCTGGA 1740  
Qy |||  
1898 TCCCGAGTGGGAGTTCGTGAAACACCCCGCCCTGTTGAAGCTGTGTTAGCTGAGTGGGA 1957  
Db |||  
1741 TCCCGAGTGGGAGTTCGTGAAACACCCCGCCCTGTTGAAGCTGTGTTAGCTGAGTGGGA 1800  
Qy |||  
1958 AGAGGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCGAGACCA 2017  
Db |||  
1801 AGAGGCCCATCTGTGGGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCGAGACCA 1860  
Qy |||  
2018 AGATCGGCAAGGCGCGCTACGTGACCGACCGGCGCGCGCAGAGAATCGTGAGCTGACCG 2077  
Db |||  
1861 AGCTGGCAAGGCGCGCTACGTGACCGACCGCGCGCGCGCAGAGAAGTGTGTGAGCATCGCGC 1920  
Qy |||  
2078 AGACCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTTGGCGCTGCGAGACAGCGGCA 2137  
Db |||  
1921 ACACCAACCAAGAGACCGAGCTGAGGCGCATCCACCTGGCGCTGCGAGACAGCGGCGC 1980  
Qy |||  
2138 GCGAGGTGAACATCTGTGACCGACAGCAGTACGCCCTGGGCGCATCATCCAGGCCAGCGCG 2197  
Db |||  
1981 TGGAGTGAACATCTGTGACCGACAGCAGTACGCCCTGGGCGCATCATCCAGGCCAGCGCG 2040  
Qy |||  
2198 ACAAGAGCGAGAGCGAGCTGGTGAACCGAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2257  
Db |||  
2041 ACAAGAGCGAGAGCGAGCTGGTGAACCGAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2100  
Qy |||  
2258 TGTACCTGAGCTGGGTGCCCGCCACAGGGCATCGGCGCGCAAGGACAGATCGACAGC 2317  
Db |||  
2101 TGTACCTGGCTGGGTGCCCGCCACAGGGGCATCGGCGCGCAAGGACAGTGGAGACAGC 2160  
Qy |||  
2318 TGTGAGCAAGGCGCATCCGCAAGGTGCTGTCTCTGGAAGGCGCATCGATGGCGGATCTCTGA 2377  
Db |||  
2161 TGTGAGCGCGCGCATCCGCAAGGTGCTGTCTCTGGAAGGCGCATCGATGGCGGATCTCTGA 2220  
Qy |||  
2378 TCTACCAAGTACATGGAAGCAGCTGTGATGGGCGAGCGCGCCCTCTAGATCGATTTAAAGC 2437  
Db |||  
2221 TCTACCAAGTACATGGAAGCAGCTGTGATGGGCGAGCGCGCCCTCTAGATCGATTTAAAGC 2280  
Qy |||  
2438 TTCCCGGGCTAGCACCGGTGAATTC 2463  
Db |||  
2281 TTCCCGGGCTAGCACCGGTGAATTC 2306

## RESULT 2

US-09-475-515-84  
; Sequence 84, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:

APPLICANT: BARNETT, Susan  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: SRIVASTAVA, Indresh  
APPLICANT: LIAN, Ying  
APPLICANT: HARTOG, Karin  
APPLICANT: LIU, Hong  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 84  
LENGTH: 2312  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: FS(-) protmod.RTopt(+)  
US-09-475-515-84

Query Match 82.2%; Score 2025.2; DB 4; Length 2312;  
Best Local Similarity 93.3%; Pred. No. 0;  
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

Qy 170 GCGGCAAGGAGGCGCACAGATGAAGACTGACGAGGCGGCGGCAACTTCTTCGCG 229  
Db 1 GCGGCGCGGAGGACACCAATGAAGATTGCACTGAGAGAGAGGCTAATTTCTTCGCG 60

Qy 230 AGGACCTGGCTTCCCGCAGGCGAGGCGCGAGTTCGCCAGGAGCAGAACCGCGCA 289  
Db 61 AGGACCTGGCTTCTGCGGCGAGGCGCGAGTTCAGGAGGAGGAGCGCGCA 120

Qy 290 ACAGCCCCCAGCGCGAGCTGCAAGGTGCGGCGG-----ACNACCCCGCAGCGAGG 343  
Db 121 ACAGCCCCCAGCGCGAGCTGCAAGGTGCGGCGGCGGAGAACAGCGCTGAGCGAGG 180

Qy 344 CCGGCGCGGAGGCGGAGGCGGCGG-----AAGTCCCGCAGATACCTCTGGGAGC 397  
Db 181 CCGGCGCGGAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

Qy 398 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 457  
Db 241 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300

Qy 458 CCGAGCAGACCGCTGCTGAGGAGATGAGCTGCGCGGCAAGTGGAGGCCAAGATGATCG 517  
Db 301 CCGAGCAGACCGCTGCTGAGGAGATGAGCTGCGCGGCAAGTGGAGGCCAAGATGATCG 360

Qy 518 CCGGATCGGCGGCTTCAAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 577  
Db 361 CCGGATCGGCGGCTTCAAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420

Qy 578 GCAAGAGCGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 637  
Db 421 GCAAGAGCGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

Qy 638 ACATGCTGACCGCTGCGGCTGCACTTGAAGTTCGCCCATCAGCGCGGCGGCGGCGG 697  
Db 481 ACCTGCTGACCGCATCGGCTGCACTTGAAGTTCGCCCATCAGCGCGGCGGCGGCGG 540

Qy 698 CCGTGAAGCTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 757  
Db 541 CCGTGAAGCTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

Qy 758 AGAGATCAAGCGGCTGACCGGCTTCTGGAGGAGATGGAGAGGAGGCGGCGGCGGCGG 817  
Db 601 AGAGATCAAGCGGCTTCTGGAGATCTGCAACCGAGATGGAGAGGAGGCGGCGGCGG 660

Qy 818 AGATCGGCGCGGAGAACCCCTTACAAACCCCGCTGTTCCCATCAAGAGAGGAGCAGCA 877  
Db 661 AGATCGGCGCGGAGAACCCCTTACAAACCCCGCTGTTCCCATCAAGAGAGGAGCAGCA 720

Qy 878 CCAAGTGGCGCAAGCTGCTGAGCTTCCGCGAGCTGAACAAGCGCACCCAGAGCTTCTGG 937  
Db 721 CCAAGTGGCGCAAGCTGCTGAGCTTCCGCGAGCTGAACAAGCGCACCCAGAGCTTCTGG 780

Qy 938 AGGTGAGCTGGGATCCCGCACCCCGCGGCTTCAAGAGAGAGAGAGAGAGAGAGAG 997  
Db 781 AGGTGAGCTGGGATCCCGCACCCCGCGGCTTCAAGAGAGAGAGAGAGAGAGAGAG 840

Qy 998 TGAAGCTGGGCGAGCGCTTCTTACGCTGCGGCGGCTTCCGAGAGAGAGAGAGAGAG 1057  
Db 841 TGAAGCTGGGCGAGCGCTTCTTACGCTGCGGCGGCTTCCGAGAGAGAGAGAGAGAG 900

Qy 1058 CTTTACCATCCCGAGCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117  
Db 901 CTTTACCATCCCGAGCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

Qy 1118 TGCCCGAGGCTGGAAGGCGAGCGGCGGCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 1177  
Db 961 TGCCCGAGGCTGGAAGGCGAGCGGCGGCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 1020

Qy 1178 AGCCCTTCCGCGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1231  
Db 1021 AGCCCTTCCGCGCGGAG 1080

Qy 1232 TGCGCAGGAGCTGGAGATCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1291  
Db 1081 TGCGCAGGAGCTGGAGATCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140

Qy 1292 TGCTGCGCTGGGCTTCAACACCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1351  
Db 1141 TGCTGCGCTGGGCTTCAACACCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

Qy 1352 GGATGGGCTACGAGTGCACCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1411  
Db 1201 GGATGGGCTACGAGTGCACCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

Qy 1412 AGGAGAGCTGAGCGTGAACAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1471  
Db 1261 AGGAGAGCTGAGCGTGAACAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320

Qy 1472 AGATCTACCCCGGCGATCAAGGTGCGCGGAGCTGTGCAAGCTGTGCGCGGCGGCGGAG 1531  
Db 1321 AGATCTACCGCGGCGATCAAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380

Qy 1532 TGAACGAGATCGTCCCTGACCGAGAGGCGGAGCTGGAGTGGCGGAGAGAGAGAGAGAG 1591  
Db 1381 TGAACGAGTGTATCCCTGACCGAGGAGGCGGAGCTGGAGTGGCGGAGAGAGAGAGAGAG 1440

Qy 1592 TCCTGCGGAGCGCTGCGAGCGGCTGTACTAGACCCCGAGAGAGAGAGAGAGAGAGAGAGAG 1651  
Db 1441 TCCTGGAAGAGCGCTGCGAGCGGCTGTACTAGACCCCGAGAGAGAGAGAGAGAGAGAGAG 1500

Qy 1652 TCCAGAGAGAGGCGGCGAGCGAGCTTACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAG 1711  
Db 1501 TCCAGAGAGAGGCGGCGAGCGAGCTTACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAG 1560

Qy 1712 TGAAGACGCGCAAGTGAAG 1771  
Db 1561 TGAAGACGCGCAAGTGAAG 1620

Qy 1772 CCGAGGCGGCTGAG 1831  
Db 1621 CCGAGGCGGCTGAG 1680

Qy 1832 TCCGCTGCGGCTTCCAG 1891  
Db 1681 TCAAGCTGCGGCTTCCAG 1740

Qy 1892 CCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGG 1951

```
Db 1741 CCTGATCCCGAGTGGAGTTCGTGAACACCCCCCTGGTGAAGCTGTGGTACCAGC 1800
Qy 1952 TGGAGAGAGCCCATCATCGGCGCCGAGACCTTCTAGCTGGAGCGGCGCCCAACCGCG 2011
Db 1801 TGGAGAGAGAGCCCATCGTGGCGCCGAGACCTTCTAGCTGGAGCGGCGCCCAACCGCG 1860
Qy 2012 AGACCAAGATCGCAAGCGCGCTACGTGACCGACCGGCGCCGAGCAAGATCGTGAGCC 2071
Db 1861 AGACCAAGCTGGGCAAGCGCGCTACGTGACCGACCGGCGCCGAGCAAGATCGTGAGCA 1920
Qy 2072 TGACCGAGACACCAACAGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTCGCAGGACA 2131
Db 1921 TGCCCGACACCAACAGAGACCGAGCTGCAGGCCATCCACCTGGCCCTCGCAGGACA 1980
Qy 2132 GCGGCGAGCGAGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGGCATCATCAGGCC 2191
Db 1981 GCGGCTTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGGCATCATCAGGCC 2040
Qy 2192 AGCCCGACAGAGCGAGCGAGCTGTGTGAACAGGCGCATCGAGCGAGCTGATCAAGAAG 2251
Db 2041 AGCCCGACAGAGCGAGCGAGCTGTGTGAACAGGCGCATCGAGCGAGCTGATCAAGAAG 2100
Qy 2252 AGAAGGTGTACCTGAGCTGGGTGCCGCCACAGGCGCATCGAGCGAGCTGATCAAGATCG 2311
Db 2101 AGAAGGTGTACCTGGCTGGGTGCCGCCACAGGCGCATCGAGCGAGCTGATCAAGATCG 2160
Qy 2312 ACAAGCTGTGAGCAAGGCGATCGGCAAGTGCTGTCTTGGAGCGGATCGATGGCGGCA 2371
Db 2161 ACAAGCTGTGAGCGCGCGGATCGGCAAGTGCTGTCTTGGAGCGGATCGATGGCGGCA 2220
Qy 2372 TCGTGATCTACAGTACAGAGCGACCTGTACGTGGGCGGCGGCGCTAGGATCGATT 2431
Db 2221 TCGTGATCTACAGTACAGAGCGACCTGTACGTGGGCGGCGGCGCTAGGATCGATT 2280
Qy 2432 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2463
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 3
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475.515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.rTopt.YMMH
US-09-475-515-83
```

```
Query Match 82.0%; Score 2019.2; DB 4; Length 2300;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;
```

```
Qy 170 GCAGCAAGAGGAGCCACCATGAAAGGACTGCAACGAGCGCCAGGCGCAACTTCTTCGCG 229
Db 1 GCAGCGCGAGAGACACCAANTGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60
Qy 230 AGACCTGGCTTTTCCCCAGGGCAAGCGCGAGTTCCCGAGTTCCTCCAGCGAGCAGAAACCGCGCCA 289
Db 61 AGACCTGGCTTTTCTGAGGGCAAGCGCGCGAGTTTCAGCAGCGAGCAGACCGCGCGCCA 120
Qy 290 ACAGCCCCCAGCGCGGAGCTCGAGGTGCGCGGG-----ACNACCCCGCAGCGAGG 343
Db 121 ACAGCCCCCAGCGCGGAGCTCGAGGTGTTGGGCGGCGAGAACAAAGCTGAGCGAGG 180
Qy 344 CCGCGCGCGAGCGCCAGGCGACCTG-----AACTTCCCCCAGATCACCTCTGTGGCAGC 397
Db 181 CCGCGCGCGAGCGCGCAGGCGACCTGTGAGCTTCAACTTCCCCCAGATCACCTCTGTGGCAGC 240
Qy 398 GCGCCCTGTGTAGCATCAAGGTGGGCGCGCAGATCAAGAGAGGCCCTGTCTGGACACCGCGCG 457
Db 241 GCGCCCTGTGTGACCATCAGGATCGGCGCGCAGCTCAAGGAGGCGCTGCTCGACACCGCGCG 300
Qy 458 CGACGACACCGTGTGTGAGGAGATGAGCTGCGCGGCAAGTGGAGGCCCAAGATGATCG 517
Db 301 CGACGACACCGTGTGTGAGGAGATGAACTGCGCGGCAAGTGGAGGCCCAAGATGATCG 360
Qy 518 GCGGCATCGCGCGCTTCATCAAGGTGCGCCAGTACGACAGATCCTGTATCGAGATCTGCG 577
Db 361 GCGGATCGGCGGCTTCATCAAGTGGCGAGTACGACAGATCCTCCGCTGGAGATCTGCG 420
Qy 578 GCAAGAGGCGCATCGGACCGTGTCTGATCGCGCGCCACCGCGGTGAACATCATCGGCGCGCA 637
Db 421 GCACAAAGGCGCATCGGACCGTGTGTGGGCGCCACCGCGGTGAACATCATCGGCGCGCA 480
Qy 638 ACATGCTGACCGAGCTGGGCTGCACTTGAACCTTCCCCCATCAGCCCATCGAGACCGTGC 697
Db 481 ACCTGCTGACCGAGTGGCTGCACCTGAACTTCCCCCATCAGCGCCCATCGAGACCGTGC 540
Qy 698 CCGTGAAGCTGAAGCGCGCATGGAACCGCGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAG 757
Db 541 CCGTGAAGCTGAAGCGCGGAGTGAACCGCGCGCGGAGTGAACCGCGCGCGGAGAGAGAGAGAG 600
Qy 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAAGAGAGAGAGAGAGAGAGAGAGAG 817
Db 601 AGAAGATCAAGGCGCTGTTGAGATCTGCAACCGAGATGGAAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy 818 AGATCGGCGCGAGAGACCCCTCAACACCGCGCGCTTCCGATCAAGAGAGAGAGAGAGAGAGAGAG 877
Db 661 AGATCGGCGCGAGAGACCCCTCAACACCGCGCGCTTCCGATCAAGAGAGAGAGAGAGAGAGAGAG 720
Qy 878 CCAAGTGGCGCAAGCTGGTGAACCTTCCGCGAGCTGAAACAGCGCACCCAGAGACTTCTGGG 937
Db 721 CCAAGTGGCGCAAGCTGGTGAACCTTCCGCGAGCTGAAACAGCGCACCCAGAGACTTCTGGG 780
Qy 938 AGGTGAGCTGGGCGATCCCCACCGCGCGCGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 997
Db 781 AGGTGAGCTGGGCGATCCCCACCGCGCGCGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Qy 998 TGAGAGTGGCGAGCGCTACTTTCAGCGTGGCCCTTGGAGCGAGGACTTCCGCAAGTACACCG 1057
Db 841 TGAGAGTGGCGAGCGCTACTTTCAGCGTGGCCCTTGGAGCAAGGACTTCCGCAAGTACACCG 900
Qy 1058 CTTTCAACATCCCGCAGCATCAACAGAGACCCCGGCGATCCGCTACAGTACAAACGCTGC 1117
Db 901 CTTTCAACATCCCGCAGCATCAACAGAGACCCCGGCGATCCGCTACAGTACAAACGCTGC 960
Qy 1118 TGCCCCAGGCTGGAAGGCGAGCCCGCAGCATCTTTCAGAGAGAGATGACCAAGATCTCTGG 1177
Db 961 TGCCCCAGGCTGGAAGGCGAGCCCGCAGCATCTTTCAGAGAGAGATGACCAAGATCTCTGG 1020
Qy 1178 AGCCCTTCCGCGCGCGCAACCGCGAGATCGTGATCTTACAGGCGCCCTGTGTGTGTGTGTGTGTGT 1237
Db 1021 AGCCCTTCCGCGCGCGCAACCGCGAGATCGTGATCTTACAGGCGCCCTGTGTGTGTGTGTGTGTGT 1080
```





QY 441 CCTGCTGGACACCGGCGCCGACACACCGTGTGGAGGAGATGATCGCTGCCCGGCAAGTG 500  
Db 1541 GCTGTCTGCACACCGCGCCGACACACCGTGTGGAGGAGATGAACTTGCCCGGCAAGTG 1600  
QY 501 GAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAGGTGCGCCAGTACGACAGAT 560  
Db 1601 GAAGCCCAAGATGATCGCGGCGATCGCGGCTTTCATCAAGGTGCGCGAGTACGACAGAT 1660  
QY 561 CTTGATGAGATCTGGCGGCAAGAGGCGCATCGGGAACGCTGTCTGATCGGCCCCCAACCCCGT 620  
Db 1661 CCCCGTGGAGATCTGGCGGCAAGAGGCGCATCGGGAACGCTGTCTGATCGGCCCCCAACCCCGT 1720  
QY 621 GAACATCATCGCGGCAAGATCTGACCCAGCTGGGCTGACCCCTGAACTTCCCATCAG 680  
Db 1721 GAACATCATCGCGGCAAGATCTGACCCAGCTGGGCTGACCCCTGAACTTCCCATCAG 1780  
QY 681 CCCCATCGAGACCGTGCCTGTAAGCTGAAAGCCCGGATGGAAGGCGCCCAAGGTGAAGCA 740  
Db 1781 CCCCATCGAGACCGTGCCTGTAAGCTGAAAGCCCGGATGGAAGGCGCCCAAGGTGAAGCA 1840  
QY 741 GTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGGAGAGATGGAGAA 800  
Db 1841 GTGGCCCTGACCGAGAGAGATCAAGGCCCTGAGGCCCTGGTGAGATCTGCAACGAGATGGAGAA 1900  
QY 801 GGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTCAACAACCCCGTGTTCGCCAT 860  
Db 1901 GGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTCAACAACCCCGTGTTCGCCAT 1960  
QY 861 CAAGAAAGAGACAGCACCAAGTGGGCGCAAGCTGGTGGATCTCCGCGAGCTGAACAAGCG 920  
Db 1961 CAAGAAAGAGACAGCACCAAGTGGGCGCAAGCTGGTGGATCTCCGCGAGCTGAACAAGCG 2020  
QY 921 CACCCAGGATCTTGGAGAGTGAGCTGGGCGATCCGCCACCCCGCGGCTGAAAGAGAA 980  
Db 2021 CACCCAGGATCTTGGAGAGTGAGCTGGGCGATCCGCCACCCCGCGGCTGAAAGAGAA 2080  
QY 981 GAAGAGGTGACCGTGTGACCTGGGCGAGCGCTACTTTCAGAGTGCCCTGACGAGGA 1040  
Db 2081 GAAGAGGTGACCGTGTGACCTGGGCGAGCGCTACTTTCAGAGTGCCCTGACGAGGA 2140  
QY 1041 CTTCCGCAAGTACACCGCTTTCACATCCCGAGCATCAACAAGAGACCCCGGCGATCCG 1100  
Db 2141 CTTCCGCAAGTACACCGCTTTCACATCCCGAGCATCAACAAGAGACCCCGGCGATCCG 2200  
QY 1101 CTAACAGTACAACTGTGCTGCCAGAGGCTGGAGGCGAGCCCGAGCATCTTTCAGAGCAG 1160  
Db 2201 CTAACAGTACAACTGTGCTGCCAGAGGCTGGAGGCGAGCCCGGCGATCTTTCAGAGCAG 2260  
QY 1161 CATGACCAAGATCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCTACCA --- 1217  
Db 2261 CATGACCAAGATCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCTACCA 2320  
QY 1218 ---GGGCCCCCTGTACGTGGGCGAGCAGCTGGAGATCGGCGAGCAGCCCGGCGCAAGATCGA 1274  
Db 2321 CATGGACGACCTGTACGTGGGCGAGCAGCTGGAGATCGGCGAGCAGCCCGGCGCAAGATCGA 2380  
QY 1275 GAGCTGCGCAAGAACCTGTGCTGGGCTTTCACACCCCGGCAAGAGACACAGAA 1334  
Db 2381 GAGCTGCGCGAGCACCTGTGCTGGGCTTTCACACCCCGGCAAGAGACACAGAA 2440  
QY 1335 GGAGCCCCCTTCTGTGTGATGGCTACGAGCTGACCCCGGCAAGGTGGAGCCGTGCAGCC 1394  
Db 2441 GGAGCCCCCTTCTGTGTGATGGCTACGAGCTGACCCCGGCAAGGTGGAGCCGTGCAGCC 2500  
QY 1395 CATCGAGCTCGCGGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAA 1454  
Db 2501 CATCATGCTCGCGGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAA 2560  
QY 1455 GCTGAATCGGCGCAGCAGATCTACCCCGGATCAAGGTGGCGGCTGTGAGCTGCT 1514  
Db 2561 GCTGAATCGGCGCAGCAGATCTACCCCGGATCTACCCCGGATCAAGGTGAAGCAGCTGTGCAAGCTGCT 2620

QY 1515 GCSCGCGCCAGAGCCCTGACCGACATCGTGCCCTGACCGAGAGGCGCGAGCTGGAGCT 1574  
Db 2621 GCSCGCGCCAGAGCCCTGACCGAGCTGATCCCTCTGACCGAGAGGCGCGAGCTGGAGCT 2680  
QY 1575 GGCCGAGAACCGGAGATCTTGCGGAGCCCTGTGCAAGGCTGTACTACGACCCAGCAA 1634  
Db 2681 GGCCGAGAACCGGAGATCTTGAGGAGCCCTGTGCAAGGCTGTACTACGACCCAGCAA 2740  
QY 1635 GGAACCTGTGGCGGCGAGATCCAGAGCAGGCGGCGACGACGAGTGGACCTTACCAGATCTACCA 1694  
Db 2741 GGAACCTGTGGCGGCGAGATCCAGAGCAGGCGGCGAGGCGGAGTGGACCTTACCAGATCTACCA 2800  
QY 1695 GGAGCCCTTCAAGAACCTGGAAGACCGGCGCAAGTACGCGCAAGATCGCACCCCGCACACAA 1754  
Db 2801 GGAGCCCTTCAAGAACCTGGAAGACCGGCGCAAGTACGCGCGCATCGCGGCGCGCACACAA 2860  
QY 1755 CGACGTGAAGACGCTGACGAGCCGTGCAAGAGATCGCCATCGGAGAGCATCTGTGATCTG 1814  
Db 2861 CGACGTGAAGACGCTGACGAGCCGTGCAAGAGATCGCCATCGGAGAGCATCTGTGATCTG 2920  
QY 1815 GGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTTGGAGACCTTGGTGGAC 1874  
Db 2921 GGCGAAGATCCCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTTGGAGACCTTGGTGGAT 2980  
QY 1875 CGACTACTGGCAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACAACCCCGCCCTGGT 1934  
Db 2981 GGAGTACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACAACCCCGCCCTGGT 3040  
QY 1935 GAAGCTGTGTGTACACGCTGGAGAGAGCCCATCATCTGGCGCGGAGACCTTCTACGTGGA 1994  
Db 3041 GAAGCTGTGTGTACACGCTGGAGAGAGCCCATCTGTGGCGCGGAGACCTTCTACGTGGA 3100  
QY 1995 CGSCGCGCCCAACCGGAGACCAAGATCGSAAAGCGCGCTAGCTGACCGACCGGGGCGG 2054  
Db 3101 CGSCGCGCCCAACCGGAGACCAAGCTGGSAAAGCGCGCTAGCTGACCGACCGGGGCGG 3160  
QY 2055 GCAGAAAGATCGTGAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCGCATCCA 2114  
Db 3161 CCAGAGGTGGTGGAGCATCGCCGACACCAACCAACAGAGACCGAGCTGCAGGCGCATCCA 3220  
QY 2115 GCTGGCCCTGACGAGACAGCGGCGAGGCTGAACATCTGTGACCGACGAGCCAGTACGCCCT 2174  
Db 3221 CTTGGCCCTGACGAGACAGCGGCGCTGGAGGTGAACATCTGTGACCGACGAGCCAGTACGCCCT 3280  
QY 2175 GGSCATCATCCAGGCGCCAGCCGACAGAGGAGCGAGCTGTGAACCAAGATCATCGA 2234  
Db 3281 GGSCATCATCCAGGCGCCAGCCGACAGAGGAGCGAGCTGTGAACCAAGATCATCGA 3340  
QY 2235 GCAGCTGATCAAGAGGAGAGGTGTACTGTGAGCTGGGTGCCCGCCCAAGGGGATCGG 2294  
Db 3341 GCAGCTGATCAAGAGGAGAGGTGTACTGTGAGCTGGGTGCCCGCCCAAGGGGATCGG 3400  
QY 2295 CGGCAACGAGCAGATCGAAGCTGGTGGAGAGGCGATCCGCAAGGTGCTGTCTCTGGA 2354  
Db 3401 CGGCAACGAGCAGATCGAAGCTGGTGGAGAGGCGATCCGCAAGGTGCTGTCTCTGGA 3460  
QY 2355 CGGCATCATGCGCGGATCTGTGATCTACCACTA 2387  
Db 3461 CGGCATCATGCGCGGATCTGTGATCTACCACTA 3493

## RESULT 5

US-09-475-515-80

; Sequence 80, Application US/09475515A

; Patent No. 6602705

; GENERAL INFORMATION:

; APPLICANT: BARNETT, Susan

; APPLICANT: ZUR MEHDE, Jan

; APPLICANT: SRIVASTAVA, Indresh

; APPLICANT: LIAN, Ying

; APPLICANT: HARTOG, Karin

; APPLICANT: LIU, Hong

; APPLICANT: GREER, Catherine

```

; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(+).proinact.RTopt.YM
US-09-475-80

Query Match      76.3%; Score 1878.8; DB 4; Length 2305;
Best Local Similarity 89.4%; Pred. No. 7.2e-312;
Matches 2061; Conservative 0; Mismatches 232; Indels 13; Gaps 3;

Qy 170 GCGGCAAGAGGCGCCACAGATGAAGGACTGCACCGAGCGCGAGCCCAACTTCTTCGCG 229
Db      |||
Qy 230 AGGACCTGGCTTCCCCCAGGCGAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCA 289
Db      |||
Qy 60 AAGATCTGGCTTCTTACAAGGGAAGGCCACGGGAATTTTCTTCAGAGCAGACGAGGCCA 119
Db      |||
Qy 290 ACAGCCGCGAGCGCGAGCTGCAGTGCGCGGGG-----ACAAACCCCGCAGCGAGG 343
Db      |||
Qy 120 ACAGCCGCGAGCGAGCTTCAGGTTTGGGAGGAGAAACAACTCCCTCTCAGAA 179
Db      |||
Qy 344 CCGGCGCGAGCGCGAGGCA-----CCCTGAATCTCCCGAGATCACTCTGGCGAGC 397
Db      |||
Qy 180 CAGAGCGGATAGACAGGAAGTGTATCTTTAACTTCTCCCTCAGATCACTCTTTGGCAAC 239
Db      |||
Qy 398 GCCCGCTGGTGAGATCAAGTGGCGCGCCAGATCAAGAGGCGCTGTGGACACCGCG 457
Db      |||
Qy 240 GACCCCTCGTCACAATGAAGATCGGGGGCAACTCAAGGAAGCGTGTGATACAGGAG 299
Db      |||
Qy 458 CCGAGCAGACCGTGTGGAGAGATGAGCCCTGCGCGGCAAGTGAAGCCCAAGATGATCG 517
Db      |||
Qy 300 CAGATGATACAGTATTAGAAAGAAATGAATTTGCCAGGAAATGGAAACCAAAATGATAG 359
Db      |||
Qy 518 GCGGCATCGCGCGCTTCATCAAGTGGCGCAGTACGACCCAGATCTGATCCAGATCTGCG 577
Db      |||
Qy 360 GGGGATCGGGGGCTTCATCAAGGTGAGGCGAGTACGACCCAGATACCTGTGAAATCTGTG 419
Db      |||
Qy 578 GCAAGAAGGCCCATCGGCACCGTGTGATCGGCGCCACCCCGTGAAACATCATCGCGCGCA 637
Db      |||
Qy 420 GACATAAGCTATAGTACAGTATTAGTAGACCTACACCTGTCAATATTGGAGAA 479
Db      |||
Qy 638 ACATGCTGACCCAGTGGCTGCAACCTGAACTTCCCGCATCAGCCCATCGAGACCGTGC 697
Db      |||
Qy 480 ATCTGTTGACCCAGATCGGCTGCACTTGAATTTCCCGCATCAGCCCTATTGAGACGGTGC 539
Db      |||
Qy 698 CCGTGAAGCTGAAGCCCGGATGAGCGCCCAAGGTGAAGCAGTGGCGCCCTGAGCGAGG 757
Db      |||
Qy 540 CCGTGAAGTGAAGCCCGGATGAGCGCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599
Db      |||
Qy 758 AGAAGATCAAGGCCCTGACCCCATCTCGAGGAGATGGAGAGGAGGCGCAAGATCACCA 817
Db      |||
Qy 600 AGAAGATCAAGGCCCTGTTGAGATCTGACCGGAGATGGAGAGGCGCAAGATCAGCA 659
Db      |||
Qy 818 AGATCGGCGCGAGAACCCCTTACCAACACCCCGCTGTTGCGCATCAAGAGAGAGCAGCA 877
Db      |||
Qy 660 AGATCGGCGCGAGAACCCCTTACCAACACCCCGCTGTTGCGCATCAAGAGAGAGCAGCA 719
Db      |||
Qy 878 CCAAGTGGCGCAAGCTGTGAGCTTCGCGAGCTGAACCAAGCGCACCCAGACTTCTGGG 937
Db      |||
Qy 720 CCAAGTGGCGCAAGCTGTGAGCTTCGCGAGCTGAACCAAGCGCACCCAGACTTCTGGG 779
Db      |||

```

```

Qy 938 AGGTGAGCTGGGCAATCCCCACCCCGCGCCCTGGAAGAAAGAGCGTGAACCGTGC 997
Db      |||
Qy 780 AGGTGAGCTGGGCAATCCCCACCCCGCGCCCTGGAAGAAAGAGCGTGAACCGTGC 839
Db      |||
Qy 998 TGGACGTGGGCGAGCGCTTCTTCAAGCTGCGCCCTGGAGCAGGACTTCCGCAAGTACACCG 1057
Db      |||
Qy 840 TGGACGTGGGCGAGCGCTTCTTCAAGCTGCGCCCTGGAGCAGGACTTCCGCAAGTACACCG 899
Db      |||
Qy 1058 CTTTACCATCCCCCAGCATCAAAACAGAGACCCCGCGCATCCGCTACAGTACACAGTGC 1117
Db      |||
Qy 900 CTTTACCATCCCCCAGCATCAAAACAGAGACCCCGCGCATCCGCTACAGTACACAGTGC 959
Db      |||
Qy 1118 TGCCCCAGAGGCTGGAAGGCGAGCCCGCAGATCTTTCCAGAGCAGCATGACCAAGATCTCTGG 1177
Db      |||
Qy 960 TGCCCCAGAGGCTGGAAGGCGAGCCCGCAGATCTTTCCAGAGCAGCATGACCAAGATCTCTGG 1019
Db      |||
Qy 1178 AGCCCTTCCGCGCGCCGCAACCCCGAGATCGTGATCTTACAGGCGCCCTGCTACGTTGGGCA 1237
Db      |||
Qy 1020 AGCCCTTCCGCGCGAGCAACCCCGAGATCGTGATCTTACAGGCGCCCTGCTACGTTGGGCA 1079
Db      |||
Qy 1238 GCGACCTGGAGATCGGCGAGCACCGCGCCAAAGATCGAGGAGCTGCGCAAGCACCTGTGTC 1297
Db      |||
Qy 1080 GCGACCTGGAGATCGGCGAGCACCGCGCCAAAGATCGAGGAGCTGCGCAGCACCTGTGTC 1139
Db      |||
Qy 1298 GCTGGGCTTTCACCAACCCCGCAAGAGCAACCAAGAGGAGCCCGCTTCTGTGGATGG 1357
Db      |||
Qy 1140 GCTGGGCTTTCACCAACCCCGCAAGAGCAACCAAGAGGAGCCCGCTTCTGTGGATGG 1199
Db      |||
Qy 1358 GCTACGAGCTGCACCCCGCAAGTGGAGCCGTGAGCCCATCGAGTGCAGGAGGAGA 1417
Db      |||
Qy 1200 GCTACGAGCTGCACCCCGCAAGTGGAGCCGTGAGCCCATCATGCTGCGCCGAGAGGACA 1259
Db      |||
Qy 1418 GCTGGAACCTGGAACGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCT 1477
Db      |||
Qy 1260 GCTGGAACCTGGAACGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCT 1319
Db      |||
Qy 1478 ACCCGCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGGCGCCAAAGGCCCTGAGCG 1537
Db      |||
Qy 1320 AGCGCGCATCAAGGTGGAAGCAGCTGTGCAAGCTGTGCGGCGCAACCAAGGCCCTGAGCG 1379
Db      |||
Qy 1538 ACATGTCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCG 1597
Db      |||
Qy 1380 AGTGTATCCCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGA 1439
Db      |||
Qy 1598 GCGAGCCGCTGCGAGCGGTGTACTACGACCCCAAGAGGAGCTGCTGGTGGCGCAGATCCAGA 1657
Db      |||
Qy 1440 AGGAGCCGCTGCGAGCGGTGTACTACGACCCCAAGAGGAGCTGCTGGTGGCGCGAGATCCAGA 1499
Db      |||
Qy 1658 AGCAGGCGCCAGACAGTGGAGCTTACAGATCTTACAGGAGCCCTTCAAGAACCTGGAAGA 1717
Db      |||
Qy 1500 AGCAGGCGCCAGGCGCAGTGGACCTTACAGATCTTACAGGAGCCCTTCAAGAACCTGGAAGA 1559
Db      |||
Qy 1718 CCGGCAAGTACGCGCAAGATGGGCAACCGCGCCACCAACAGAGCTGAGCAGTGAACCGAGG 1777
Db      |||
Qy 1560 CCGGCAAGTACGCGCGCGATGCGCGCGCCCAACCAAGAGCAGTGAACCGAGG 1619
Db      |||
Qy 1778 CCGTGAAGAGATCGCCATGAGAGAGCTCGTGATCTGGGGCAAGACCCCGCAAGTTTCGCGC 1837
Db      |||
Qy 1620 CCGTGAAGAGATGAGCAGCAGAGAGATCGTGATCTGGGGCAAGATCCCCAAGTTTCAAGC 1679
Db      |||
Qy 1838 TGCCCATCCAGAGAGGAGACCTGGGAGACCTGGTGGAGCCGACTACTGGCAGGCGCACCTGGA 1897
Db      |||
Qy 1680 TGCCCATCCAGAGAGGAGACCTGGGAGGCGCTGGTGGATGGAGTACTGGCAGGCGCACCTGGA 1739
Db      |||
Qy 1898 TCCCGAGTGGAGTTCGTGAAACACCCCGCCCTGGTGAAGCTGTGGTACAGCTGGAGA 1957
Db      |||
Qy 1740 TCCCGAGTGGAGTTCGTGAAACACCCCGCCCTGGTGAAGCTGTGGTGAAGCTGTGGTGAAG 1799
Db      |||
Qy 1958 AGGAGCCCATCATCGGCGCGAGAGCTTCTACGTGAGAGCGCGCGCCCAACCGCGAGACCA 2017
Db      |||
Qy 1800 AGGAGCCCATCATCGGCGCGAGAGCTTCTACGTGAGAGCGCGCGCCCAACCGCGAGACCA 1859
Db      |||

```

QY 2018 AGATCGCAAGCGCGCTACGTGACCGACCGGGCGGCGAGAAATCGTGAGCTGACCG 2077  
|||  
DB 1860 AGCTGGCCAGCGCGCTACGTGACCGACCGGGCGGCGAGAAAGGTGGTGAGCATCGCCG 1919  
|||  
QY 2078 AGACCAACCAACAGAGACCGAGCTGCGAGGCCATCCAGCTGGCCCTGCGAGGACAGCGGCA 2137  
|||  
DB 1920 ACACCAACCAACAGAGACCGAGCTGCGAGGCCATCCACCTGGCCCTGCGAGGACAGCGGCC 1979  
|||  
QY 2138 GCGAGGTGAACATCGTGACCGACGAGCCAGTACCGCCCTGGCGCATCCAGGCCCGAGCCCG 2197  
|||  
DB 1980 TGGAGGTGAACATCGTGACCGACGAGCCAGTACCGCCCTGGCGCATCCAGGCCCGAGCCCG 2039  
|||  
QY 2198 ACAAGAGCGAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAAGAGAAAGG 2257  
|||  
DB 2040 ACAAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAAGAGAAAGG 2099  
|||  
QY 2258 TGTACTGTAGCTGGGTGCCGCCACAAAGGCGATCGGGCGCAACGAGCAGATCGACAAGC 2317  
|||  
DB 2100 TGTACTGTGGCTGGGTGCCGCCACAAAGGCGATCGGGCGCAACGAGCAGGTGGAACAAGC 2159  
|||  
QY 2318 TGGTGAGCAAGGCATCCGCAAGGTGCTGTTCTGGACGGCATCGATGGCGGCATCGTGA 2377  
|||  
DB 2160 TGGTGAGCGCGGCATCCGCAAGGTGCTGTTCTGAAACGGCATCGATGGCGGCATCGTGA 2219  
|||  
QY 2378 TCTACAGTACATGGACGACCTGTACGTGGCGAGCGCGGCCCTAGGATCGATTAAAGC 2437  
|||  
DB 2220 TCTACAGTACATGGACGACCTGTACGTGGCGAGCGCGGCCCTAGGATCGATTAAAGC 2279  
|||  
QY 2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463  
|||  
DB 2280 TTCCCGGGGCTAGCACCGGTGAATTC 2305  
|||

## RESULT 6

US-09-475-515-81  
; Sequence 81, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 81  
; LENGTH: 2299  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(+).proinact.RTopt.YMMW

## US-09-475-515-81

Query Match 75.2%; Score 1852; DB 4; Length 2299;  
Best Local Similarity 89.0%; Pred. No. 2,7e-307;  
Matches 2052; Conservative 0; Mismatches 235; Indels 19; Gaps 4;

QY 170 GCGCAAGGAGGCGCCACCACTGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCGCG 229  
|||  
DB 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGACAGGCTTAATTTTT-AGGG 59  
|||  
QY 230 AGGACCTGGCTTTCCCGCAGGCGAAGCCCGCGAGTTCCCGCAGCAGAGCAACCGGCCA 289  
|||

DB 60 AAGATCTGGCTTCTTACAAGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCCAGAGCCA 119  
|||  
QY 290 ACAGCCCCACCCAGCCGCGAGCTGCAGGTGCGCGCG-ACAAACCCCGCAGCGAGG 343  
|||  
DB 120 ACAGCCCCACCCAGAGAGAGCTTCAGGTTTGGGAGGAGAAACAATCTCTCTCAGAG 179  
|||  
QY 344 CCGGCGCCGAGCGCCAGGGCA-----CCGTGAATCTTCCCGCAGATCACCTCTGTGGCAGC 397  
|||  
DB 180 CAGGAGCGATAGACAAGGAATGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239  
|||  
QY 398 GCCCCTCTGTGAGCATCAAGTGGGCGGCCAGATCAAGGAGGCCCTGCTGGACACCGGCG 457  
|||  
DB 240 GACCCCTCTGTCAATAAAGGATCGGGGGCACTCAAGGAAGCGCTGCTCGATACAGGAG 299  
|||  
QY 458 CCAGACGACCCGTCTGGAGGAGATGAGCTGCCGCAAGTGAAGGCCCAAGATGATCG 517  
|||  
DB 300 CAGATGATACAGTATTAGAAAGAAATGAATTTGTCAGGAAATGAAACCAAAATGATAG 359  
|||  
QY 518 GCGGCATCGCGCGCTTCATCAAGGTGCGCCAGTACGACACAGATCCTGATCGAGATCTGCG 577  
|||  
DB 360 GGGGATCGGGGCTTCATCAAGGTGAGGCGAGTACGACACAGATACCTGTAGAAATCTGTG 419  
|||  
QY 578 GCAAGAGGCGCATCGGCACGTCTGATGGGCCCGACCCCGGTGAACATCATCGGCCGCA 637  
|||  
DB 420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAAATTTGGAAGAA 479  
|||  
QY 638 ACATGTCGACCCAGCTGGGCTGCACCTGAACTTCCCGCATCAGCCCATCGAGACCGTGC 697  
|||  
DB 480 ATCTGTTGACCCAGATCGGCTGCACCTTGAATTTCCCGCATCAGCCCTATTGAGACGGTGC 539  
|||  
QY 698 CCGTGAAGCTGAAGCCCGCATGGAAGCGGCCCGCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 757  
|||  
DB 540 CCGTGAAGTTGAAGCCGGGATGGAAGCGGCCCGCAAGGTCAAGCAATGGCCATTGACCGAGG 599  
|||  
QY 758 AGAAGATCAAGGCCCTGACCGCATCTGCGAGGAGATGGAAGAGGAGGCGCAAGATCAACCA 817  
|||  
DB 600 AGAAGATCAAGGCCCTGTTGGAGATCTGCACCGAGATGGAGAGAGGAGGCGCAAGATCAAGCA 659  
|||  
QY 818 AGATCGGCGCCGAGAACCCCTACAACACCCCGCTTCCGCATCAAGAAGAAGAGAGCA 877  
|||  
DB 660 AGATCGGCGCCGAGAACCCCTACAACACCCCGCTTCCGCATCAAGAAGAAGAGAGCA 719  
|||  
QY 878 CCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGAAACAGCGCAACCCAGGACTTCTGGG 937  
|||  
DB 720 CCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGAAACAGCGCAACCCAGGACTTCTGGG 779  
|||  
QY 938 AGTGAGCTGGGCATCCCGACCCCGCGCTGGAAGAGAGAGAGAGGAGTGACCGTGC 997  
|||  
DB 780 AGGTGAGCTGGGCATCCCGACCCCGCGCTGGAAGAGAGAGAGAGGAGTGACCGTGC 839  
|||  
QY 998 TGGACGTGGCGAGCGCTACTTTCAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCG 1057  
|||  
DB 840 TGGACGTGGCGAGCGCTACTTTCAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCG 899  
|||  
QY 1058 CTTTCACCATCCCGCAGCATCAACAAAGAGACCCCGCGCATCCGCTACAGTACAACTGC 1117  
|||  
DB 900 CTTTCACCATCCCGCAGCATCAACAAAGAGACCCCGCGCATCCGCTACAGTACAACTGC 959  
|||  
QY 1118 TGCCCGAGGCTGGAGAGGCGAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCTGG 1177  
|||  
DB 960 TGCCCGAGGCTGGAGAGGCGAGCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCTCG 1019  
|||  
QY 1178 AGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATCTACAGGCCCGCCCTGCTACGTGGGCA 1237  
|||  
DB 1020 AGCCCTTCGCGAAGAGAGACCCCGCATCTGATCTTACAGGCCCGCCCTGTACGTGGGCA 1079  
|||  
QY 1238 GCGACCTTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCAAGCACTGCTGC 1297  
|||  
DB 1080 GCGACCTTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCGCAGCACTGCTGC 1139  
|||  
QY 1298 GCTGGGGCTTCACACCCCGCAAGAGCACAAGAGGAGGCCCTTCTCTGTGGATGG 1357  
|||  
DB 1140 GCTGGGGCTTCACACCCCGCAAGAGCACAAGAGGAGGCCCTTCTCTGCCAT-- 1197  
|||

```
Qy 1358 GCTACGAGCTGCACCCCGGCAAGTGGACCGGTGACGCCCATCGAGTGCCTCCGAGAGGAGA 1417
Db 1198 ----CGAGCTGCACCCCGGCAAGTGGACCGGTGACGCCCATCATGCTGCCCGGAGGAGCA 1253
Qy 1418 GCTGCGCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1477
Db 1254 GCTGAGCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1313
Qy 1478 ACCCGGCGATCAAGGTGGCGGAGCTGTGCAAGCTGTGCGGGGCCCAAGGCCCTGACCG 1537
Db 1314 AGCGCGGCGATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGGGGCCCAAGGCCCTGACCG 1373
Qy 1538 ACATGTCCTCTGACCGAGAGCGGAGCTGGAGCTGGGCCGAGAACCGCGAGATCTCTGC 1597
Db 1374 AGGTGATCCCCCTGACCGAGAGGCGGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGA 1433
Qy 1598 GCGAGCCCGTGCACGGGTGTACTACGACCCAGCAAGGACCTGTGTGCCGAGATCCAGA 1657
Db 1434 AGGAGCCCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGTGTGCCGAGATCCAGA 1493
Qy 1658 AGCAGGCGCCAGCAGCTGGACCTTACCAGATCTTACGAGGAGCCCTTCAAGAACCTGAAGA 1717
Db 1494 AGCAGGCGCCAGGCGGAGCTGGACCTTACCAGATCTTACGAGGAGCCCTTCAAGAACCTGAAGA 1553
Qy 1718 CCGCAAGTACGCCAAGATGGGCAACCGGCCACACCAAGCAGCTGAAGAGCTGACCGAGG 1777
Db 1554 CCGCAAGTACGCCCGCATGCGCGCGGCCACACCAAGCAGCTGAAGAGCTGACCGAGG 1613
Qy 1778 CCGTGCAAGAGATCGCCATGAGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCC 1837
Db 1614 CCGTGCAAGAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1673
Qy 1838 TGCCCATTCGAAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGA 1897
Db 1674 TGCCCATTCGAAGAGGAGACCTGGGAGGCTGTGTGATGAGTACTTGGCAGGCCACCTGGA 1733
Qy 1898 TCCCGGAGTGGAGTTCTGTGAACACCCGCCCTTGGTGAAGCTGTGTGTAACAGCTGGAGA 1957
Db 1734 TCCCGGAGTGGAGTTCTGTGAACACCCGCCCTTGGTGAAGCTGTGTGTAACAGCTGGAGA 1793
Qy 1958 AGGAGCCCATCATCGGCGCGGAGACCTTCTAGCTGGAGCGGCGGCCCAACCGCGAGACCA 2017
Db 1794 AGGAGCCCATCATCGGCGCGGAGACCTTCTAGCTGGAGCGGCGGCCCAACCGCGAGACCA 1853
Qy 2018 AGATCGGCAAGGCGCGGCTTACGTGACCGACCGGCGCGGCGGAGAGATCGTGAGCCTGACCG 2077
Db 1854 AGCTGGGCAAGGCGCGCTTACGTGACCGACCGGCGCGGCGGAGAGAGGTGTTGAGCATCGCG 1913
Qy 2078 AGACCAACCAAGAGACCGAGTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137
Db 1914 ACACCAACCAAGAGACCGAGTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGCGGCG 1973
Qy 2138 GCGAGGTGACANTCGTGACCGAGCGGAGTACGCGCTGGGCGCATCATCCAGGCCAGCCCG 2197
Db 1974 TGGAGGTGAACATCGTGACCGAGCGGAGTACGCGCTGGGCGCATCATCCAGGCCAGGCCCG 2033
Qy 2198 ACAAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGG 2257
Db 2034 ACAAGAGCGAGCGAGCTGTGTGAGCCAGATCATCGAGCAGCTGATCAAGAGAGAGAGG 2093
Qy 2258 TGTAACCTGAGTGGGTGCCCGCCCAAGGCGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db 2094 TGTACCTGGCTGGGTGCCCGCCCAAGGCGCATCGGCGGCAACGAGCAGGTGGAACAAGC 2153
Qy 2318 TGGTACCAAGGGGCGATCGGCAAGTGTGTTCCTGGAGCGGATCGATCGGGCGCATCTGA 2377
Db 2154 TGGTACGCGCGGCGATCGGCAAGTGTGTTCCTGAAACGCGCATCGATCGGGCGCATCTGA 2213
Qy 2378 TCTACCAAGTACATGAGCAGCTGTACTGGGCGAGCGGCGCTAGGATCGATTAAAGC 2437
Db 2214 TCTACCAAGTACATGAGCAGCTGTACTGGGCGAGCGGCGCTAGGATCGATTAAAGC 2273
```

```
Qy 2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db 2274 TTCCCGGGGCTAGCACCGGTGAATTC 2299

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-synGp - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

Query Match 67.1%; Score 1651.8; DB 4; Length 4307;
Best Local Similarity 82.5%; Pred. No. 4.2e-273;
Matches 1939; Conservative 0; Mismatches 391; Indels 19; Gaps 4;

Qy 33 GGCACACAGCGCCAAATCTCTGATGACGCGCAGCAACTTCAAGGGCCCCCAAGCGCATCAT 92
Db 1110 GACCAACTCCGCTACCATCATGATGACGCGCGGCAACTTTTCGGAACCAACGCAAGATCGT 1169
Qy 93 CAAGTGTCTCAACTGCGCAAGGAGGCGCACATCGCCGCAACTGCGCGCCGCCCGCAA 152
Db 1170 CAAGTGTCTTCACTGTGCAAGAGAGGCGCACAGCCGCAACTGCGAGGGCCCTTAGGA 1229
Qy 153 GAAGGGTGTCTGGAAGTGC CGGCAAGAGGCGCACACAGATGAAGACTGCACCGAGCGCA 212
Db 1230 AAGGGTGTCTGGAATGCGCAAGAGGCGCACACAGATGAAGACTGTACGAGAGACA 1289
Qy 213 GGCACAACTTCTTCGCGAGGACCTGGCTTCCCGAGGGCAAGGCGCGAGTTCGCCAG 272
Db 1290 GGTCAA-TTTTTTAGGGAAGATCTGGCTTCTCTACAAGGGAAGGCCAGGGAATTTCTTC 1348
Qy 273 CGAGCAGAACCGCGCCAAACAGCCACAGCGCGAGCTGCAGGT-----GCGCGGCGA 326
Db 1349 AGAGCAGACCGCGCCAAACAGCCACCGCGCGAGCTTCAGGTCTGGGGTCCGGAACA 1408
Qy 327 CAACCCCGCAGCGAGGCGCGCGCGAGCGCGCAGGCA-----CCCTGAATTTCCCGCA 380
Db 1409 CAATCCCGCTCCGAAGCAGGAGCGCGCGCGAGCGCGAGCGGTGTCTTCAACTTCCCTCA 1468
Qy 381 GATCACCCTGTGCGAGCGCCCTGTGTGAGCATCAAGTGGCGCGCCAGATCAAGAGGC 440
Db 1469 GGTCAACCTTTGGCAGCGACCCCTCGTCAACCATCAAGATCGGGGGGCGAGCTCAAGAGGC 1528
Qy 441 CTTGCTGAGACACCGCGCGCGAGCAGCACCGTGTCTGGAGAGATGAGCTGCCCGGCAAGTG 500
Db 1529 TCTCTGGACACCGGAGCAGACGACACCGTGTCTGGAGAGATGTCTGTCCAGCGCGCTG 1588
Qy 501 GAAGCCCCAAGATGATCGCGCGCATTCGCGCGCTTCAAGGTGGCCAGTACGACCAAGAT 560
Db 1589 GAAGCCGAAGATGATCGGGGGAATCGCGGTTTTTCATCAAGGTGCGCCAGTATGACAGAT 1648
Qy 561 CTTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCGTGTCTGATCGGCCCGCACCCCGT 620
Db 1649 CCTCATCGAAATCTGCGGCCACAGGCGTATCGGTACCGTGTCTGTGGGGCCCGCACACCGT 1708
Qy 621 GAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCGCATCAG 680
Db 1709 CAACATCATCGGAGCGCAACCTGTTGACGCGAGATCGGTGCACTGAACTTCCCGCATTAG 1768
```

QY 681 CCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGGTGAAGCA 740  
 Db |||||  
 QY 1769 CCCTATCGAGACCGTACCGGTGAAGCTGAAGCCCGGATGAGCGGCCGAGGTCAAGCA 1828  
 Db |||||  
 QY 741 GTGGCCCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAA 800  
 Db |||||  
 QY 1829 ATGGCCATTGACAGAGAGAGATCAAGGCACTGGTGGAGATTGTCACAGAGATGGA 1888  
 Db |||||  
 QY 801 GGAGGGAAGATCACCAAGATCGGCCCGGAGAACCCCTCAACACACCCCGTGTTCGCCAT 860  
 Db |||||  
 QY 1889 GGAAGGGAATCTCCAAGATTGGGCTTGAGAACCCGTAACAACCGCGGTGTTCGCAAT 1948  
 Db |||||  
 QY 861 CAAGAAGAAGGACGACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920  
 Db |||||  
 QY 1949 CAAGAAGAAGGATCGACGAANAATGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 2008  
 Db |||||  
 QY 921 CACCCAGGACTTCTGGAGAGTGCAGCTGGGCACTCCCCACCCCGCGGCTGAAGAAGAA 980  
 Db |||||  
 QY 2009 CACGCAAGACTTCTGGGAGGTTCAAGCTGGGCACTCCCGCACCCCGCAGGGCTGAAGAAGAA 2068  
 Db |||||  
 QY 981 GAAGAGGTGACCGTGTGAGCTGGGCGAGCGCTACTTCAGAGTGGCCCTGACGAGGA 1040  
 Db |||||  
 QY 2069 GAAATCCGTGACCGTACTGAGATGGGTGATGCTACTTCTCCGTTCCCTCGGACGAGA 2128  
 Db |||||  
 QY 1041 CTTCCGCAAGTACACCGCTTTCACCATCCCGACATCAACAAGAGACCCCGGCACTCCG 1100  
 Db |||||  
 QY 2129 CTTTCAGAGTACACTGCTTCAATCCCTTCGATCAACAGAGACACCGGGATTG 2188  
 Db |||||  
 QY 1101 CTTACAGTACAACTGCTGCCCCAGGGCTGGAAGGCGAGCCCCAGCATCTTCCAGAGCAG 1160  
 Db |||||  
 QY 2189 ATATCAGTACAACTGCTGCCCCAGGGCTGGAAGGCTTCCCGCAATCTTCCAGAGTAG 2248  
 Db |||||  
 QY 1161 CATGACAGATCCTTGAGGCCCTTCGCGGCCGGAACCCCGAGATCGTGNATCTACCA --- 1217  
 Db |||||  
 QY 2249 CATGACCAAAATCTCGAGGCTTTCGCAAAAGAGACCCCGACATCGTCACTATCAGTA 2308  
 Db |||||  
 QY 1218 ---GGCCCCCTGTAGTGGGCGAGCACTGGAGATCGGCGACGCCCGCCCAAGATCGA 1274  
 Db |||||  
 QY 2309 CATGATGACTTGTAGTGGGCTCTGATCTAGAGATAGGCGACGACCGACCAAGATCGA 2368  
 Db |||||  
 QY 1275 GGAGCTGCGCAAGCACCTGCTGGGCTGGGGCTTTCACACCCCGCAAGAAGACCCAGAA 1334  
 Db |||||  
 QY 2369 GGAGCTGCGCGACGACCTGTTGAGTGGGACTGACACACCCGCAAGAAGCACCCAGAA 2428  
 Db |||||  
 QY 1335 GGAGCCCCCTTCTGTGATGGGCTACGAGCTGCAACCCCGCAAGTGGACCGTGCAGCC 1394  
 Db |||||  
 QY 2429 GGAGCCCTCCCTTCTGTGATGGGTTACGAGCTGCAACCCCTGACAAATGGACCGTGCAGCC 2488  
 Db |||||  
 QY 1395 CATCGAGCTGCCGAGAGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGTGGGCAA 1454  
 Db |||||  
 QY 2489 TATCGTGTCCGAGAGAGAGCTGGAATGTCACAGCATACAGAGCTGTGGGGAA 2548  
 Db |||||  
 QY 1455 GCTGAATGGGCGCAGCAGATCTACCCCGCATCAAGGTGCGGCGAGCTGTGCAAGCTGCT 1514  
 Db |||||  
 QY 2549 GTTGAATGGGCGCAGTCAGATTACCCAGGATTAAGTGTAGGCGAGCTGTGCAAACTCCT 2608  
 Db |||||  
 QY 1515 GCGCGGCGCAAGGCCCTGACGACATCTGTGCCCTTGACCGAGAGGCGCGAGCTGAGCT 1574  
 Db |||||  
 QY 2609 CCGCGGAACCAAGGCACTCAAGAGGTGATCCCTTAACCGAGAGGCGCGAGCTCGAAT 2668  
 Db |||||  
 QY 1575 GCGCGGAGAACCGGAGATCTGCGCGAGCCGCTGCAAGGCTGTACTACGACCCCGAGCAA 1634  
 Db |||||  
 QY 2669 GGCAAGAAACCGAGAGATCTTAAGAGGCCCGCTGCAAGGCTGTACTATGACCCCTCCAA 2728  
 Db |||||  
 QY 1635 GGAACCTGGTGGCGGAGATCCAGAAGCAGGCGCCACGACCACTGAGCTTACAGATCTACCA 1694  
 Db |||||  
 QY 2729 GGAACCTGATCGCGAGATCCAGAAGCAGGCGCAAGGCGCACTGACCTATCAGATTTACCA 2788  
 Db |||||  
 QY 1695 GGAACCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGTGGGACCGCGCCACCA 1754  
 Db |||||  
 QY 2789 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCCGATGAGGGGTGCCACACTAA 2848  
 Db |||||

QY 1755 CGACGTGAAGCAGCTACCCGAGCCCGTGCAGAAATCGGCATCGAGAGCATCTGTATCTG 1814  
 Db |||||  
 QY 2849 CGAGCTCAAGCAGCTGACCGAGCCCGTGCAGAAATCAACCAAGCAATCTGTATCTG 2908  
 Db |||||  
 QY 1815 GGGCAAGACCCCAAGTTCCGCTCGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGAC 1874  
 Db |||||  
 QY 2909 GGGAAAGACTCTCAAGTTCAAGCTGCCATCCAGAAAGAAACCTGGGAAACCTGGTGAC 2968  
 Db |||||  
 QY 1875 CGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCCCTGGT 1934  
 Db |||||  
 QY 2969 AGAGTATTGGCAGGCCACCTGGATCTCTGAGTGGGAGTTCTGTCAACACCCCTCCCTGGT 3028  
 Db |||||  
 QY 1935 GAAGCTGTGTGTACACAGCTGGAGAGGAGCCCATCATCTCGCGCCGAGACCTTTACGTGGA 1994  
 Db |||||  
 QY 3029 GAAGCTGTGTGTGTACAGCTGGAGAGGAGCCCATAGTGGGCGCCGAAACCTTTCTACGTGGA 3088  
 Db |||||  
 QY 1995 CGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGGCTAGCTGACCGACCGGGCGG 2054  
 Db |||||  
 QY 3089 TGGGGCGGCTTAAACAGGAGACTAAGCTGGGCAAGCGGATACGTCACTAACCGGGGAG 3148  
 Db |||||  
 QY 2055 GCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCA 2114  
 Db |||||  
 QY 3149 ACAGAGGTTGTCAACCTCACTGACACCAACCAAGAGACTGAGCTGCAGGCCATTTA 3208  
 Db |||||  
 QY 2115 GCTGGCCCTGACGAGCAGCGGCGAGAGGTGAACATCTGACCGACAGCAGTACGCCCT 2174  
 Db |||||  
 QY 3209 CCTCGCTTTGCGAGACTCGGGCTGGAGGTGAACATCTGTGACAGACTCTCAGTATGCCCT 3268  
 Db |||||  
 QY 2175 GGGCATCATCCAGGCCCGCCGACAAAGAGCGAGAGCTGTTGNAACAGATCATCGA 2234  
 Db |||||  
 QY 3269 GGGCATCATTCAGGCCCGCCGACAGCAGAGTGAAGTCCGAGCTGTTCAATCAGATCATCGA 3328  
 Db |||||  
 QY 2235 GCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTCGCCGCCCAAGGGCATCGG 2294  
 Db |||||  
 QY 3329 GCAGCTGATCAAGAAGAAAGGTCTATCTGGCTGGGTACCGCCCAAGAGGCAATGG 3388  
 Db |||||  
 QY 2295 CGGCAACGAGCAGATCGACAAAGCTGGTGAAGGGCATCCGCAAGGTGCTGTTCTCTGGA 2354  
 Db |||||  
 QY 3389 CGGCAATGAGCAGGTGACAAAGCTGCTCGGCTGGCATCAGGAAGTGTCTATTCTCTGGA 3448  
 Db |||||  
 QY 2355 CGGCATCGA 2363  
 Db |||||  
 QY 3449 TGGCATCGA 3457  
 Db |||||

RESULT 8

US-09-936-572-2  
 ; Sequence 2, Application US/09936572  
 ; Patent No. 6783981  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UDEN. MARK  
 ; APPLICANT: MITROPHANOUS, KYRIACOS  
 ; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
 ; FILE REFERENCE: 07883/0137  
 ; CURRENT APPLICATION NUMBER: US/09/936, 572  
 ; CURRENT FILING DATE: 2001-12-11  
 ; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
 ; PRIOR FILING DATE: 2000-03-17  
 ; PRIOR APPLICATION NUMBER: GB 9906177.2  
 ; PRIOR FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 4307  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:  
 ; OTHER INFORMATION: gagpol-SYNGp-codon optimised gagpol sequence  
 US-09-936-572-2

Query Match 66.3%; Score 1632.6; DB 4; Length 4307;  
 Best Local Similarity 82.0%; Pred. No. 8e-270;

Matches 1927; Conservative 0; Mismatches 403; Indels 19; Gaps 4;			
Qy	33	GGCCACAGCGCCCAACATCTCTGATGCGCGCAGCAACTTCAAGGGCCCCCAAGCGCATCAT	92
Db	1110	GACCAACTCCGCTACCATCATGATGATGAGCGCGGCAACTTTCGGAAACCAACGCAAGATCGT	1169
Qy	93	CAAGTGTCTTCACTGCGCGCAAGAGGGCCACATCGCCGCGCAACTGCGCGCCCGCCGCA	152
Db	1170	CAAGTGTCTTCACTGTGCGCAAGAGGGCCACATCGCCGCGCAACTGCGCGCCCGCCGCA	1229
Qy	153	GAAGGGTGTCTGGAAGTGGCGCAAGAGGGCCACCATGATGAAGACTGCAACCGAGCGCCA	212
Db	1230	AAAGGGTGTCTGGAAGTGGCGCAAGAGGGCCACCATGATGAAGACTGTAAGAGAGACA	1289
Qy	213	GGCCAACTTCTTCGCGAGGACCTGGCTTTCCCGCAGGCAAGGGCCGCGAGTTTCCCGAG	272
Db	1290	GGCTAA-TTTTATAGGGAAGATCTGGCTTCTTACAAGGGAAGGCCAGGGAATTTTCTTC	1348
Qy	273	GGAGCAGAACCGCGCCCAACAGCCCCACAGCGCGGAGCTGAGTGGCGG-----CGA	326
Db	1349	AGAGCAGACAGAGCCCAACAGCCCCACAGAGAGAGCTTCAAGTCTGGGGTAGAGACAA	1408
Qy	327	CAACCCCGCAGGAGGCGCGCGCCGAGCGCCAGGCA-----CCCTGAACTTTCCCCCA	380
Db	1409	CAACTCCCCCTCAGAACAGAGCCGATAGAACAGGAATCTGTATCTTTAACTTCCCTCA	1468
Qy	381	GATCACCTCTGGCAGCGCCCTCTGTGTGAGCATCAAGTGGCGGCGCAGATCAAGGAGGC	440
Db	1469	GATCACTTCTTGGCAACGACCCCTCGTCAATAAAGATAGGGGGGCGAGCTCAAGGAGGC	1528
Qy	441	CTGTCTGGACACCGCGCCGACGACACCTGTGTGGAGGAGATGAGCCCTGCGCGGAAGTG	500
Db	1529	TCTCTGGACACCGGAGCAGACACACCTGTGTGGAGGAGATGTGTGTCAGGCGCGCTG	1588
Qy	501	GAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAAGTGTGGCGCAGTACGACACAGAT	560
Db	1589	GAAGCCCAAGATGATCGCGCGCAATCGCGCGTTCATCAAGTGTGGCGCAGTATGACACAGAT	1648
Qy	561	CTGTATCGAGATCTCGCGCAAGAGGCGCATCGGCACCGTGTGATCGGCGCCCGCCCGCT	620
Db	1649	CCTCATCGAATCTGCGCGCCACAGGCTATCGGTACCGTCTGTGGCGCCCGCACACCGCT	1708
Qy	621	GAACATCATCGCGCGCAACATGTGACCCAGTGGGTGCACTTGAACCTTCCCGCATCAG	680
Db	1709	CAACATCATCGCGCGCAACCTGTGTGACGAGATCGGTGTCACGCTGAACTTCCCGCATAG	1768
Qy	681	CCCATTCGACACCGTCCCGTGAAGTGAAGCCCGCATCGGCGCCCGCCCGAGGTGAGCA	740
Db	1769	CCCTATCGACACCGTACCGGTGAAGTGAAGCCCGGATGGAAGCGCCCGAGGTCAAGCA	1828
Qy	741	GTGGCCCTTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGGAGAA	800
Db	1829	ATGGCCATTGACAGAGAGAGAGATCAAGGCACTGGTGGAGATTGCAACAGATGGAATA	1888
Qy	801	GGAGGGCAAGATCACAAGATCGCGCCCGGAAACCCCTACAACAACCCCGTGTTCGCGCAT	860
Db	1889	GGAAAGGAAATCTCCAAGATTGGGCTGTGAAGACCCGTTACAACAAGCGCGGTGTTCGCAAT	1948
Qy	861	CAAGAGAGACAGACACAGTGGCGCAAGTGGCGCAAGTGGTGGACTTCCGCGAGTGAACAGCG	920
Db	1949	CAAGAGAGAGAGATCGACGAAATGGGCGCAAGTGGTGGACTTCCGCGAGTGAACAGCG	2008
Qy	921	CACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGCCCGCCCGCGCTGGAAGAGAA	980
Db	2009	CAGCGAAGACTTCTGGGAGGTTCAGCTGGGCGATCCCGCCCGCGAGGCTGGAAGAGAA	2068
Qy	981	GAAGAGCGTACCGTGTGACGCTGGCGCAGCGCTACTTCAGCGTGCCTCTGACGAGGA	1040
Db	2069	GAATCCGTGACCGTACTGATGTGGGTGATGCTACTTCTCCGTTCCCTCTGACGAGAA	2128
Qy	1041	CTTCCGCAAGTACACCGCTTACCATCCCGAGGATCAACAAGAGACCCCGCGCATCCG	1100
Db	2129	CTTCAGGAAGTACACTGCTTCAATCCCTTCGATCAACAAGAGAGACACCGGGGATTCCG	2188

Qy	1101	CTACCAAGTACAACGCTGCTGCCAGGGCTGGAAGGGCAGCGCCCGAGCATCTTTCAGAGCAG	1160
Db	2189	ATATCAGTACAACGCTGCTGCCAGGGCTGGAAGGGCTCTCCGCAATCTTCCAGAGTAG	2248
Qy	1161	CATGACCAAGATCCTGGAGCCCTTCCGCGCCGCGCAACCCCGAGATCGTGATCTTACCA---	1217
Db	2249	CATGACCAAAATCCTGGAGCCCTTCCGCAAAACAGAACCCCGACATCTATCATAGTA	2308
Qy	1218	---GGCCCCCTCTAGCTGGGCGAGCGACTGGAGATCGGCGAGCACCGCGCCAAAGATCGA	1274
Db	2309	CATGAGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGCGAGCACCGCACCAAGATCGA	2368
Qy	1275	GGAGCTGCGCAAGCACTGTGCTGGCTGGGGCTTCAACACCCCGCAAGAGACCAAGAA	1334
Db	2369	GGAGCTGCGCGCAGCACCTGTGAGTGGGGACTGACCACACCCCGCAAGAGACCAAGAA	2428
Qy	1335	GGAGCCCCCTTCTGTGATGGGCTACGAGCTGCAACCCCGCAAGTGGAGCCGTGCGAGCC	1394
Db	2429	GGAGCCCCCTTCTGTGATGGGTTACGAGCTGCAACCCCGCAAGTGGAGCCGTGCGAGCC	2488
Qy	1395	CATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAA	1454
Db	2489	TATCGTGTCTGCCAGAGAAAGACAGCTGGAATGTCAAGCAATACAGAAAGCTGGTGGGAA	2548
Qy	1455	GCTGAATCTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCGAGCTGTGCAAGCTGCT	1514
Db	2549	GTTGAATCTGGGCCAGTCAGATTTACCCAGGGATTAAGTGTGAGGCAGCTGTGCAAACTCT	2608
Qy	1515	GCGCGGCGCAAGCCCTTGACACATCTGTGCCCTTGACCGAGGAGCGCGAGCTGGAGCT	1574
Db	2609	CCGCGAAACCAAGGCACTCAAGAGGTGATCCCCCTTAACCGAGGAGCGCGAGCTCGAAT	2668
Qy	1575	GGCGGAGAACCGGAGATCTGCGGAGCCCGTGCAGCGGTGTACTACGACCCCGCAAGAA	1634
Db	2669	GGCAGAAACCGAGAGATCTTAAGAGCCCGTGCAGCGGTGTACTATGACCCCTCAA	2728
Qy	1635	GGACCTGTGTGGCGCAGATCCAGAAAGCAGGGCCACGACCACTGAGCACTTACAGATCTACCA	1694
Db	2729	GGACCTGTATCGCGCAGATCCAGAAAGCAGGGGCAAGGCGACCTATCAGATTTACCA	2788
Qy	1695	GGAGCCCTTCAAGAACCTGAGACCGGCAAGTAGCCCAAGATGGGCACCGCCACACAA	1754
Db	2789	GGAGCCCTTCAAGAACCTGAGACCGGCAAGTAGCCCGAGTGGGGGTGCCACACTAA	2848
Qy	1755	CGAGCTGAAGCAGCTGACCGGAGCCGTGCAAGAGATCGCCATGGAGAGCATCGTGATCTG	1814
Db	2849	CGAGCTGAAGCAGCTGACCGGAGCCGTGCAAGAGATCACCCCGAAAGCATCGTGATCTG	2908
Qy	1815	GGCAAGACCCCAAGTTCCGCTCCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGAC	1874
Db	2909	GGCAAGACCTCTAAGTTCAAGCTGCCATCCAGAAAGAAACCTGGGAAACCTGGTGGAC	2968
Qy	1875	CGACTATGCGCAGGCCCTGTGATCCCGGATGGGAGTTGCTGAAACACCCCGCCCTCGGT	1934
Db	2969	AGAGTATTTGGCAGGCCACCTGATCTCTGAGTGGGAGTTCTGTCACACCCCTCCCTCGGT	3028
Qy	1935	GAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGA	1994
Db	3029	GAAGCTGTGTATCAGCTGGAGAGGAGCCCATAGTGGGCGCGAAACCTTCTACGTGGA	3088
Qy	1995	CGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGGCGGCTACGTGACCGACCGGGGCGG	2054
Db	3089	TGGGCGCGCTAACAGGAGACTAAGCTGGGCAAGCGGATACGTCACTAAACCGGGGCGG	3148
Qy	2055	GCAGAGATGCTGAGCGCTGACCGAGACCAACCAAGAGAGACCGAGCTGCGAGGCGCATCCA	2114
Db	3149	ACAGAGGTTGTCACTTCACTGACACCAACCAAGAGACTGAGCTGCGAGGCGCATTTA	3208
Qy	2115	GCTGGCCCTGCGAGCAGCGGAGGAGGTAACATCTGTGACCGAGACCGAGCTACGCCCT	2174
Db	3209	CCTCGCTTTGACGAGACTCGGGCCCTGGAGGTGAACATCTGTGACAGACTCTCAGTATGCCCT	3268



```
QY 2175 GGCATCATCATCAGGCCAGCCGACAGAGCGAGCGAGCTGGTGAACAGATCATCGA 2234
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3269 GGGCATCATTCAGACCCAGCCAGACACAGATGAGTCCGAGCTGGTCAATCAGATCATCGA 3328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2235 GCAGCTGATCAAGAAGGAGAGGTGTACTCTGAGCTGGTGCCCGCCCAACAAGGCGATCGG 2294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3329 GCAGCTGATCAAGAAGGAGAGGTGTACTCTGGCTGGGTACCCGCCCAACAAGGCGATTGG 3388
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2295 CGGCAACGACGATGACGAAGCTGGTGACAGAGGCGATCCGCAAGTGTCTTCTCGGA 2354
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3389 CGGCAATGACAGCTGACGAAGCTGGTCTCGGCTGGCATCAGGAAGTGTCTATTCTCTGGA 3448
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2355 CGGCATCGA 2363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3449 TGGCATCGA 3457
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-09-936-572-14
; Sequence 14, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936, 572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 4327
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP4-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with 20bp of the leader
; OTHER INFORMATION: sequence of HIV-1
US-09-936-572-14

Query Match 66.0%; Score 1624.6; DB 4; Length 4327;
Best Local Similarity 81.8%; Pred. No. 1.9e-268;
Matches 1922; Conservative 0; Mismatches 408; Indels 19; Gaps 4;

QY 33 GGCACACAGCGCCAAACATCTGATGACGCGCAGCAACTTCAAGGGGCCCAAGCGCATCAT 92
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1130 GACCAACTCGCTACCATCATGATGACGCGGGCNACTTTCGGNAACCAAGCGAAGATCGT 1189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 93 CAAGTGTCTCAACTGCGGCAAGAGGGGCCACATGCGCCGCAACTGCGCGCCCCCGCCAA 152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1190 CAAGTGTCTCAACTGTGGCAAGAGGGGCACACAGCCGCAACTGACAGGGCCCTTAGGA 1249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 153 GAAGGGTGTGTGAAGTGGCGGCAAGGGGCCACACAGATGAAGACTGACCCGAGCGCCA 212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1250 AAAGGGCTGTGGAAATGTGGAAAGGAAGGACACCAATGAAAGATTGTACTAGAGACA 1309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 213 GGCCAACTTCTCCGCGAGACCTGGCTTCCCGCAGGCAAGCGCGAGTTCGCCAG 272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1310 GGCTAA-TTTTATAGGAAGATCTGGCTTCCCAAGGAGGACCGGGAATTTTCTTC 1368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 273 CGAGAGAAACCGGCCAACAGCCCCACAGCCGCGAGCTGCAGGTGCGGG-----CGA 326
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1369 AGAGCAGACAGAGCCAAACAGCCCCACACAGAGAGCTTCAGTTTGGGGAAGAGACAA 1428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 327 CAACCCCCGAGGAGCGCGCGCCGAGCCGAGCCGAGGGA-----CCCTGAACTTCCCCCA 380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1429 CAATCCCTCTCAGAAGCAGGAGCCCGATAGACAAAGGAACCTGTATCTCTTTAGCTTCCCTCA 1488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 381 GATCACCTCTGTCAGCGGCCCTGGTGAGCATCAAGGTGGCGGCCAGATCAAGGAGGC 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 1489 GATCATCTTTGGCAGCGACCCCTCGTCAATAAAGATAGGGGGCAGCTCAAGGAGGC 1548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 441 CTGTCTGGACACCGCGCCGACGACACCGTGTGGAGGAGATGAGCTGCCCGCGCAAGTG 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1549 TCTCTCTGGACACCGGAGCAGACGACACCGTGTGGAGGAGATGTCGTTCGACGCGCGCTG 1608
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 501 GAAGCCCAAGATGATCGGCGGCATCGCGGCTTTCATCAAGGTGCGCCAGTACGACCAAGAT 560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1609 GAAGCCGAAGATGATCGGGGAATCGCGGTTTCATCAAGGTGCGCCAGTATGACCAAGAT 1668
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 561 CTTGATCGAGATCTGCGGCAAGAGCCATCGGACCCGTCTGATCGGCCCCACCCCGCT 620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1669 CCTCATCGAAATCTGCGGCCCAAGGCTATCGGTACCGTGTGGTGGGCCCCACACCCGT 1728
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 621 GAACATCATCGCGCCGCAACATGCTGACCCAGCTGGGCTGCGCCCTGAACTTCCCATCAG 680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1729 CAAATCATCGGACGCAACCTGTTGACGAGATCGGTTGACGCTGAACTTCCCATTAG 1788
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 681 CCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGGCCCGCCCAAGGTGAAGCA 740
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1789 CCTATCGAGACGCTACCGGTGAAGCTGAAGCCCGGATGGACGGCCCGGAGGTCAAGCA 1848
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 741 GTGSCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGATGGAGAA 800
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1849 ATGCCCATTTGACAGAGAGAGATCAAGGCCTGTTGGAGATTTGCACAGAGATGGAAA 1908
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 801 GGAGGCAAGATCACCAGATCGGCCCGGAGAACCTTCAACACACCCCGTGTTCGCCAT 860
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1909 GGAAGGGGAAAATCTCCAAGATTTGGGCTTGAGAACCCGTACAAACACCCCGTGTTCGCAAT 1968
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCG 920
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1969 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCG 2028
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 921 CACCCAGGACTTCTGGGAGGTGAGCTGGGCATCCCCCACCCCGCCGCTTGAAGAAGAA 980
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2029 CACGCAAGACTTCTGGGAGTTTCAGCTGGGCATCCCCGACCCCGCAGGGGTGAAGAAGAA 2088
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 981 GAAGAGGTGACCGTCTGGAAGTGGGCGACGCTACTTTCAGCGTGCCTTCGACGAGGA 1040
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2089 GAAATCCGTGACCGTACTGGAATGGGTGATGCTACTTCTCGTCTCCCTTCGACGAGGA 2148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1041 CTTCCGCAAGTACACCGCTTTCACCATCCCGACATCAACAAAGAGACCCCGGCATCCG 1100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2149 CTTCAAGAGTACACTGCTTCAATCCCTTCGATCAACAAAGAGACACCGGGGATTCG 2208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1101 CTAACAGTACAACTGCTGCCCGAGGGTGGAAAGGCGAGCCCGCAGCATCTTCCAGAGCAG 1160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2209 ATATCAGTACAACTGCTGCCCGAGGGCTGGAAGGGCTCTCCCGCAATCTTCCAGAGTAG 2268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1161 CATGACCAAGATCTTGGAGCCCTTTCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2269 CATGACCAAAATCTGAGCGCTTTCGCAAAAGAACCCCGCATCTGTCATCTATCAGTA 2328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1218 ---GGCCCCCTCTAGCTGGGCGAGCAGCTGGAGATCGGCGCAGCAGCCGCGCAAGATCGA 1274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2329 CATGATGACTGTGAGTGGGCTCTGATCTAGAGATAGGCGACGACCGCACCAAGATCGA 2388
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1275 GGAGCTGCCCAAGCACCTGCTGCGCTGGGGCTTTCACACCCCGCAAGAAGCAACCAAGAA 1334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2389 GGAGCTGGCGCAGCACCTGTTGAGGTGGGACTGACACACCCCGCAAGAAGCACCAGAA 2448
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1335 GGAGCCCCCTTCTGATGGGTACGAGTGGCTACCCCGACAGTGGACCGTGGAGCC 1394
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2449 GGAGCCTCCCTCTCTGAGTGGGTTCAGAGCTGCAACCCCTGACAAATGGACCGTGGAGCC 2508
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1395 CATCGAGCTGCCCGAGAAAGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGTGTGGGCAA 1454
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2509 TATCGTGTGCCAGAGAAAGACAGCTGGAATGTCAACGACATACAGAGCTGTGTGGGAA 2568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1455 GCTGAACTGGGCCAGCAGATCTTACCCCGCATCAAGGTGCGCGCAGCTGTGCAAGCTGT 1514
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 2569 GTTGAATGGGCGAGTCAGATTATCCCAAGGANTTAAGGTGAGGCGAGCTGTGCAAACTCCT 2628  
Qy 1515 GCGGGGCCCAAGGCCCTGACCGACATCGTGCCTCGACCGAGGAGGCGGAGCTGAGCT 1574  
Db 2629 CCGCGAACCAGGCACTCACAGAGGTGATCCCTCAACCGAGGAGGCGGAGCTGCACT 2688  
Qy 1575 GCGCGAGAACCGGAGATCCTGCGGAGCGCGTGCACGCGGTGTACTAGACCCGAGAA 1634  
Db 2689 GGCAGAAACCGAGAGATCCTAAAGAGCCGCTGCACGCGGTGTACTATGACCCCTCCAA 2748  
Qy 1635 GGACTGTGTGCGGAGATCCAGAGAGCGGCGCACGACAGTGGACCTACAGATCTACCA 1694  
Db 2749 GGAATGATGCGCGAGATCCAGAGAGCGGCGAGCGGCGGAGCTTATCATGATTTACCA 2808  
Qy 1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGGCCACACAA 1754  
Db 2809 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCGGATGAGGGGTGCCACACTAA 2868  
Qy 1755 CGAGTGAAGAGCTGACCGAGGCGGTGCAGAGATCGCCATGAGAGCATCGTGTATCTG 1814  
Db 2869 CGAGCTCAAGCAGCTGACCGAGGCGGTGCAGAGATCACACCGAAGACATCGTGTATCTG 2928  
Qy 1815 GGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGAGACCTGGGAGACCTGTGGAC 1874  
Db 2929 GGGAAAGACTCTAAGTTCAAGCTGCCATCCAGAGAGAACCTGGGAAACCTGTGGAC 2988  
Qy 1875 CGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGGT 1934  
Db 2989 AGATTAATGGCAGGCCACCTGGATCTCTGAGTGGGAGTTCTGTCAACACCCCTCCCTGGT 3048  
Qy 1935 GAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGCGCCGAGACCTTCTACGTGGA 1994  
Db 3049 GAAGCTGTGTACAGCTGGAGAGAGGCCCATATAGTGGCGCCGCAAACTTCTACGTGGA 3108  
Qy 1995 CGGCGCCGCAACCGCAGACCAAGATCGGCAAGCGCGCTACCTGACCGACCGGCGCG 2054  
Db 3109 TGGGGCCGCTAACAGGGAGACTAAGCTGGGCAAGCGGATACCTGACTAAACCGGGGAG 3168  
Qy 2055 GCAGAAGATCTGAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAAGGCCATCCA 2114  
Db 3169 ACAGAAGTTGTACCCCTCACTGACACCAACCAACAGAGACTGAGCTGCAAGGCCATTA 3228  
Qy 2115 GCTGGCCCTGAGAGACGCGGAGGAGGTGAACATGCTGACCGACGACGAGTACGCCCT 2174  
Db 3229 CCTCGCTTTCAGGACTCGGCGCTGAGGTGAACATCGTGCACAGACTCTCAGTATGCCCT 3288  
Qy 2175 GGGCATCATCCAGSCCGACGACAGAGCGAGCGAGCTGTGTGAACACAGATCATCGA 2234  
Db 3289 GGGCATCATTCAGSCCGACGACAGAGCGAGTGTGTGAGTGTGTGATCATCATCATCGA 3348  
Qy 2235 GCAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGCGCCCGCCACAGGCGCATCGG 2294  
Db 3349 GCAGCTGATCAAGAGGAGAGGTGTACTGCGCTGGGTACCGCCCGCCACAGGCGCATGG 3408  
Qy 2295 CGGCAAGCAGAGATCGCAAGAGCTGGTGAACAGGCGATCCGCAAGGTGTCTTCTGGA 2354  
Db 3409 CGGCAATGAGCAGTCCAGAGCTGGTCTCGGCTGGCATCAGGAAGGTGTCTTCTCTGGA 3468  
Qy 2355 CGGCATCGA 2363  
Db 3469 TGGCATCGA 3477

## RESULT 10

US-09-936-572-13

; Sequence 13, Application US/09936572  
; Patent No. 6783981  
; GENERAL INFORMATION:  
; APPLICANT: UDEN, MARK  
; APPLICANT: MITROPHANOUS, KYRIACOS  
; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
; FILE REFERENCE: 078863/0137  
; CURRENT APPLICATION NUMBER: US/09/936,572

; CURRENT FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: GB 9906177.2  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 4353  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: pSVNGP3-codon  
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence from  
; OTHER INFORMATION: the major splice donor  
US-09-936-572-13

Query Match 66.0%; Score 1624.6; DB 4; Length 4353;

Best Local Similarity 81.8%; Pred. No. 1.9e+268;

Matches 1922; Conservative 0; Mismatches 408; Indels 19; Gaps 4;

Qy 33 GGCACACGAGCGCCCAACATCCTGATGCGCGCAGCAAACTTCAAGGGCCCAAGCGCATCAT 92  
Db 1156 GACCAACTCCGCTACCATCATGATGCGCGCGGCAACTTTGGAACCAACGCAAGATCGT 1215  
Qy 93 CAAGTGTCTTCAACTGCGCAAGGAGGCGCAATCGCCCGCAACTGCCCGCCCCCGCAA 152  
Db 1216 CAAGTGTCTTCAACTGCGCAAGGAGGCGCAACAGCGCGCAACTGCGCGCCCCCTAGGAA 1275  
Qy 153 GAAGGGCTGTGGAAGTGTGCGCAAGGAGGCGCAACAGATGAAGACTGCACGAGCGCA 212  
Db 1276 AAAGGGCTGTGGAAGTGTGGAAGGAGGACACCAATGAAGATTTGACTGAGAGACA 1335  
Qy 213 GGCCAACTTCTTCGCGAGGACCTGGCTTCCCGAGGCAAGGCGCGAGTTCCCGAG 272  
Db 1336 GGCTAA-TTTTTAAAGGAAGATCTGGCTTCCCAAGGGAAGGCCAGGAATTTCTTC 1394  
Qy 273 CGAGCAGAACCGCGCCCAACAGCCCCACAGCGCGAGCTGCAGGTGCGCGG-----CGA 326  
Db 1395 AGAGCAGACAGAGCCCAACAGCCCCACAGAAAGAGAGCTTCAGGTTTGGGGAAGAGACA 1454  
Qy 327 CAACCCCGGAGCGAGCGCGCGCGCGAGCGCGCAAGGCA-----CCCTGAACCTTCCCA 380  
Db 1455 CAATCCCTCTCAGAAGCAGGAGCGGATAGACAAGGAACCTGATCTCTTTAGCTTCCCTCA 1514  
Qy 381 GATCACCTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGCGCGCCAGATCAAGGAGGC 440  
Db 1515 GATCACCTTTGGCAGCGACCCCTCGTCAATTAAGATAGGGGGGCGAGCTCAAGGAGGC 1574  
Qy 441 CTTGCTGGACACCGCGCCGACGACACCGTGTCTGGAGGAGATGAGCTTCCCGCGCAAGTG 500  
Db 1575 TCTCTGGACACCGGAGCAGACGACACCGTGTGGAGGAGATGCTGTTGCCAGGCGCGTG 1634  
Qy 501 GAAGCCCAAGATGATCGCGCGCATCGCGGCTTATCAAGGTGCGCCAGTACGACCAAT 560  
Db 1635 GAAGCCCAAGATGATCGCGGGAATCGGCGGTTTATCAAGGTGCGCCAGTATGACCAAT 1694  
Qy 561 CCTGATCGAGATCTGCGCAAGAGGCGCATCGGCAACGCTGTGATCGGCGCCACCGCCGT 620  
Db 1695 CCTCATCGAATCTGCGGCGCAACAGGCTATCGGTACCGTGTGTTGGCGCCCGCACCCGT 1754  
Qy 621 GAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGCAACCTTGAACCTTCCCGCATCAG 680  
Db 1755 CAACATCATCGGACGCAACCTGTTGACGCGAGATCGTGTGCAACGCTGAACCTTCCCGCATTAG 1814  
Qy 681 CCCCATCGAGACCGTGGCGCGCATGAAGCGCGGATGAAGCGCGCCCAAGGTGAAGCA 740  
Db 1815 CCCTATCGAGACGCTACCGGTGAAGCTGAAGCCCGGGATGAGACGCGCCCGAAGGTCAAGCA 1874  
Qy 741 GTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGGAGAA 800  
Db 1875 ATGGCCATTGACAGAGAGAGATCAAGGGCACTGTTGGAGATTTGCAAGAGATGGAGAA 1934

Qy	801	GGAGGGCAAGATCATCAAGATCGCCCGCAGAACCCCTACAAACCCCGTGTTCGCAT	860
Db	1935	GGAAAGGGAATCTCCAAGATTGGCGCTGAGAACCGGTACAAACGCGGTGTTCGAAT	1994
Qy	861	CAAGAAGAAGACAGCAACCAAGTCGCGCAGCTGGTGACTTCCGCGAGCTGAACAAGCG	920
Db	1995	CAAGAAGAAGACTTCGACGAATGGCGAAGCTGGTGACTTCCGCGAGCTGAACAAGCG	2054
Qy	921	CACCCAGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCGCCGCTGAAGAAGAA	980
Db	2055	CACCCAGACTTCTGGGAGGTTCAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAA	2114
Qy	981	GAAGAGCGTGACCGTGTGTGGAGTGGGCGACGCTACTTTCAGCGTGCCCTGACGAGGA	1040
Db	2115	GAATTCGTGACCGTACTGGATGTGGGTGATGCTACTTCTCCGTTCCCTCGACGAAGA	2174
Qy	1041	CTTCCGCAAGTACACCGCTTCAACATCCCAGCATCAACAAACGAGACCCCGCATCCG	1100
Db	2175	CTTCAGGAAGTACACTCCCTTCACAATCCCTTCGATCAACAAACGAGACACCGGGATTCG	2234
Qy	1101	CTACCACTCAACCGTGTGCCCGCAGGCTTGAAGGGCAGGCCAGCATCTTCCAGAGCAG	1160
Db	2235	ATATCAGTACAACTGTGTGCCCGCAGGCTTGAAGGGTCTTCCGCGCATCTTCCAGAGTAG	2294
Qy	1161	CATGACCAAGATCTCGAGGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCA	1217
Db	2295	CATGACCAAAATCTCGAGCGCTTCCGCAAAACAGAACCCCGCATCTCTATCATCAGTA	2354
Qy	1218	---GGCCCCCTGTACTGTGGCAGCGACTTGGAGATCGGCCAGCACCGCGCCAGATCGA	1274
Db	2355	CATGGATGACTTGTACTGTGGGCTCTGATCTAGAGTAGGGCAGCACCGCACCAAGATCTGA	2414
Qy	1275	GGAGCTCGGCAAGCACTCTGCTCGCTCGGGCTTCAACACCCCGCACAAAGACACACAGAA	1334
Db	2415	GGAGCTCGGCAAGCACTCTTGAAGTGGGACTGACCAACCCCGACAAGAGCACACAGAA	2474
Qy	1335	GGAGCCCCCTTCTCTGTGGATGGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCA	1394
Db	2475	GGAGCCTCCCTTCTCTGGATGGGTTACGAGCTGCACCCCTGACAAATGGACCCGTGCA	2534
Qy	1395	CATCGAGCTGCCGAGAAAGAGAGCTGGAACGCTGGAACGCAATCCGAAAGCTGGTGGCAA	1454
Db	2535	TATCGTCTGCCACAGAAAGACAGCTGGACTGTCAACGCAATACAGAAAGCTGGTGGGAA	2594
Qy	1455	GCTCAACTGGSCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGCT	1514
Db	2595	GTTGACTGGGCACTCAGATTTACCCAGGATTTAGGTGAGGAGCTGTGCAAACTCTCT	2654
Qy	1515	GCAGCGGCGCAAGCCCTGCACGACATCTGTCGCCCTGACCGAGAGGCGCGAGCTGGAGCT	1574
Db	2655	CCGCGGAACCAAGSCATCAAGAGGTGATCCCCCTTAACCGAGAGGCGCGAGCTCGAACT	2714
Qy	1575	GGCCGAGAACCGCGAGATCTCTGCGGAGCCCGGTGCAACGGCGTGTACTACGACCCGAGCAA	1634
Db	2715	GGCAGAAAAACCGAGATCTCTAAAGAGAGCCGCTGCACGGCGGTGTACTATGACCCCTCAA	2774
Qy	1635	GGACCTGGTGGCCAGATCCAGAGCAGGGGCCAGCAGCTGGACCTTACAGATCTTACCA	1694
Db	2775	GGACCTGATCGCCGAGATCCAGAGCAGGGGCCAGGCGCAGTGGACCTTATCAGATTTACCA	2834
Qy	1695	GGAGCCCTTCAAGAACTGTAAGACCGGCAAGTAGCCCAAGATGCACACCGCCACACCAA	1754
Db	2835	GGAGCCCTTCAAGAACTGAAGACCGGCAAGTAGCCCGGATGAGGGGTGCCCACTAA	2894
Qy	1755	CGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCAATGAGAGCATCTGTGATCTG	1814
Db	2895	CGAGCTCAAGCAGCTGACCGAGGCGGTGCAAGAGATCACCAACCGAAAGCATCTGTGATCTG	2954
Qy	1815	GGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTTGGGAGACCTTGTGGAC	1874
Db	2955	GGGAAAGACTCTTAAGTTCAAGCTTGCCCATCTCAGAAAGAACTTGGGAAACCTTGTGGAC	3014
Qy	1875	CGACTACTGCGAGGCCACCTGGATTCGCCGAGTGGGAGTTGCTGTAACACCCCGCCCTGGT	1934

[illegible]

```

RESULT 11
US-936-572-12
; Sequence 12, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; CURRENT REFERENCE: 078883/0137
; FURTHER APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patcin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: optimised HIV-1 gagpro
US-936-572-12

```

	Query Match	66.08;	Score 1624.6;	DB 4;	Length 4642;
	Best Local Similarity	81.8;	Pred. No. 1.9e-268;		
	Matches 1922;	Conservative	0;	Mismatches 408;	Indels 19; Gaps 4;
Qy	33	GGCCACAGCGCCCAACATCTCTGATGTCAGCGCAGCAACTTCAAGGGGCCCAAGCGCATCAT	92		
Db	1445	GACCAATCCGCTACCATCATGATGTCAGCGCGCAACTTTCGAAACCAACGCAAGATCGT	1504		
Qy	93	CAAGTGCTTCAACTGCGGCAGAGAGGGCCACATCGCCCGCAATGCGCGCCCGCCCGCAA	152		
Db	1505	CAAGTGCTTCAACTGTGTGCAAGAGGGCACAAGCCCGCACTGCAAGGCGCCCTAGGAA	1564		

Qy	153	GAAGGGCTGTGGAAAGTCGGCGAAGAGGGCCACAGATGAAGAACTGCAACCGAGCCCA	212
Db	1565	AAAGGGCTGTTCGAAAATGTGAAAAGGAGGACACAAATGAAGATTTGTACTGAGAGACA	1624
Qy	213	GGCCAACTTCTTCGCGAGGACCTGGCCCTCCCCCGAGGGCAAGCGCCGCGAGTTCGCCAG	272
Db	1625	GGCTAA- TTTTITTAGGGHAGATCTGGCCCTTCCCAAGGGHAGGCCAGGGAAATTTCTTC	1683
Qy	273	CGAGCAGAACCGCGCCAAACAGCCCCACAGCGCGAGCTGCAGGTGGCGCG- ----CGA	326
Db	1684	AGAGCAGACAGAGCCAAACAGCCCCACAGAAAGAGAGCTTCAGGTTTGGGGAGAGACAA	1743
Qy	327	CAACCCCGCAGCAGGCGCGCGCGAGCGCCAGGGCA- ----CCTGAACTTTCGCCCA	380
Db	1744	CAACTCCCTCTCAGAAGCAGGAGCGGATAGACAAGAGAACTGTATCTCTTTAGCTTCCCTCA	1803
Qy	381	GATCACCTGTGGCAGCGCCCTCTGTGTAGCATCAAGGTGGCGCGCCAGATCAAGGAGGC	440
Db	1804	GATCACTCTTTTGGCAGCACCCCTCTGTCAAAATAAGATAGGGGGGAGCTCAAGGAGGC	1863
Qy	441	CCTGCTGSAACCGCGCGCAGCAGCACCGTGTCTGGAGGAGTAGCCTGCCCGCAAGTG	500
Db	1864	TCTCTTGACACCGGAGCAGACGACACCGTGTCTGGAGGAGATGCTTGCCAGGCCGCTG	1923
Qy	501	GAAGCCCAAGATGATCGGCGGCATCGGGGCTTCATCAAGGTGGCGCAGTACGACAGAT	560
Db	1924	GAAGCCCAAGATGATCGGGGAAATCGGCGGTTTCATCAAGGTGCGCCAGTATGACCAAGT	1983
Qy	561	CCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGTCTGATCGGCCCAACCCCGCT	620
Db	1984	CCTCATCGAATCTGCGGCCACAGGGTATCGGTACCGTGTCTGTGGGCCCAACCCCGT	2043
Qy	621	GAAATCATCGGCCCGCAACATGCTGAACCACTGGGTGACACCTGAACTTCGCCATCAG	680
Db	2044	CAACATCATCGAACGCAACCTGTTGACGCGAGATCGGTTGCACTGAACTTCGCCATTAG	2103
Qy	681	CCCCATCGAGACCGTGGCCGTGAAGCTGAGCCCGGCATGAGCGGCCCAAGGTGAGCA	740
Db	2104	CCCTATCGAGACGCTACCGGTGAAGCTGAAGCCCGGATGAGACGGCCCGGAAGGTCAAGCA	2163
Qy	741	GTGGCCCTGTACCAGGAGAAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA	800
Db	2164	ATGGCCATTGACAGAGGAGAGATCAAGGCACCTGTTGAGATTTGACAGAGATGGAAA	2223
Qy	801	GGAGGCAAGATACCAAGATGCGCCCGAGAACCCCTTACAACACCCCGTGTTCGCCAT	860
Db	2224	GGAAAGGAAAATCTCCAAGATTGGCCCTTGAGAACCCGCTACAACACGCGGTGTTCGCAAT	2283
Qy	861	CNAGAAAGGACAGCAACAGTGGCGCAGCTGTTGGAATTCGCGAGCTGACACAGCG	920
Db	2284	CNAGAAAGGAGACTCGACGAATAGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCG	2343
Qy	921	CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCCGCTGAAGAAGAA	980
Db	2344	CACGCAAGACTTCTGGGAGGTTCAAGCTGGGCATCCCGCACCCCGCAGGCTGAAGAGAA	2403
Qy	981	GAAAGAGGTGACCGTGTGACGTGGGGGAGCGCTTACTTCAGCGTGGCCCTTGACGAGGA	1040
Db	2404	GAATTCGCTGACCGTACTGGAATGTGGGTGTATGCCCTACTTCTCCGTTCCCTTGAGCAAGA	2463
Qy	1041	CTTCGCGAAGTACACCGCTTACCATCCCCAGCATCAACAGAGACCCCCCGCATCCG	1100
Db	2464	CTTCAGGAAGTACTGCTTCTCAAAATCCCTTCGATCAACACGAGAACCGGGGATTCG	2523
Qy	1101	CTACCACTACAACTGTGCCCCAGGCTGSAAGGGCAGCCCCACAGTCTTCCTCAGAGCAG	1160
Db	2524	ATATCAGTACAACTGTGCTCCCCAGGCTGSAAGGCTCTCCCGCAATCTTCCAGAGTAG	2583
Qy	1161	CATGACCAAGATCTCTGAGCCCTTCCGGCCCGCAACCCCGAGATCGTGATCTACCA- ---	1217
Db	2584	CATGACCAAAATCTCTGAGCGCTTTCGCAAAACAGAACCCCGCATCGTCACTCATCAGTA	2643

QY	1218	---	GGCCCCCTGTACTGTGGCGAGCACTGGAGATCGCGCAGCA	CGCGCCCAAGATCGA	1274
DB	2644	CATGGATGACTTTGTACGTGGGCTCTGATCTAGAGATAGGGCAGCA	CGCACCAAGATCGA	2703	
QY	1275	GGAGCTGCGAAAGCACCTCTGCGCTGGGCTTCACCA	CCCCCGACAAGAAGCACAGAA	1334	
DB	2704	GGAGCTGCGCAGCACTGTGTAGGTGGGACTGACCA	CCCCCGACAAGAAGCACAGAA	2763	
QY	1335	GGAGCCCCCTTCTCTGTGAGTGGCTACGAGCTGC	ACCCCGACAAGTGGACCGTGCAGCC	1394	
DB	2764	GGAGCCCTCCCTTCTCTGTGATGGTTACGAGCTGC	ACCCCGACAAGTGGACCGTGCAGCC	2823	
QY	1395	CATCGAGCTGCCGAGAAGGAGAGCTGGACCGTGA	ACGATCATCAGAGCTGTGGGCA	1454	
DB	2824	TATCGTCTGTCAGAGAAAGACAGCTGGACTGT	CAACGACATACAGAAAGCTGTGGGAA	2883	
QY	1455	GCTGAACCTGGGCGAGCAGATCTACCCCGGCAT	CAGAGTGGCGCACGCTGTGCAAGCTGCT	1514	
DB	2884	GTTGAACCTGGGCCAGTCAGATTTACCCAGGGAT	TAAAGTGGAGGAGCTGTGCAAACTCCT	2943	
QY	1515	GCGCGGCGCAAGGCCCTGACCGACATCTGTGCC	CTGACCGCAGGAGCGCGAGCTGGAGCT	1574	
DB	2944	CCGCGAAACCAAGCACTCAGAGGTGATCCCTAA	CCGAGGAGGCCGAGCTCGA	3003	
QY	1575	GCCCGAAGACGGGAGATCTTCGCGAGCCGCTGC	ACCGGCGTGTACTGACCCCCCAGCAA	1634	
DB	3004	GGCAGAAAAACGAGAGATCTTAAAGGAGCCGT	GTGCA	3063	
QY	1635	GGAGCTGGTGGCGAGATCCAGAGCAGGCGCAC	AGCAGTGGACCTACAGATCTACCA	1694	
DB	3064	GGAGCTGATCGCGAGATCCAGAAAGCAGGGCA	AGGCGCAGTGGACCTTACAGATTTACCA	3123	
QY	1695	GGAGCCCTTCAAGAACTGTAAGACCGGCAAGT	ACGGCAAGATCGCACCGCCCAACCAA	1754	
DB	3124	GGAGCCCTTCAAGAACTGTAAGACCGGCAAGT	ACGGCAAGTGGAGGGTGCCACACTAA	3183	
QY	1755	CGAGCTGAAGAGCTGACCGAGCCGCTGCAGAA	AGATCGCAATGGAGCATCTGTGATCTG	1814	
DB	3184	CGAGCTCAAGCAGCTGACCGCAGGCGGTGC	AGAAAGATCACCAACGAAAGCATCTGTATCTG	3243	
QY	1815	GGCGAGACCCCAAGTTCCGCTCGCCATCCAG	AGAGGAGACTTGGAGACCTGTGGACCTGTGGAC	1874	
DB	3244	GGGAAAGACTCTTAAAGTTCAAGTTGCGCCAT	CTCAGAAAGAACTTGGGAACTTGGTGGAC	3303	
QY	1875	CGACTACTGCGAGCCACCTTGGATCCCGAGT	GGGAGTTGCTGAACACCCCCCCTGGT	1934	
DB	3304	AGAGTATTGCGAGGCCACTTGGATTCCTG	AGTGGGAGTTGCTGAACACCCCCCCTGGT	3363	
QY	1935	GAAGCTGTGTGATCCAGCTGGAGAAGGAGCC	ATCATCTGGCGCGAGACCTTTCTACGTGA	1994	
DB	3364	GAAGCTGTGTGATCCAGCTGGAGAAGGAGCC	ATCATCTGGCGCGAGACCTTTCTACGTGA	3423	
QY	1995	CGGCGCGCCAAACGGGAGACCAAGATTCGG	CAAGGCGGGCTAGTGACCGACCGGGGCGG	2054	
DB	3424	TGGGGCCCGCTAAACAGGGAGACTAAGCT	TGGGCAAAAGCCGATGTCATTAACCGGGGCGAG	3483	
QY	2055	GCAGAAATCGTGAGCTGACCGAGACCAAC	CAACGAGAGACCGAGCTGCAGGCCATCCA	2114	
DB	3484	ACAGAGGTTGTCACTCTACTGACCAACCA	ACCAAGAGACTGAGCTGCAGGCCATTTA	3543	
QY	2115	GCTGGCCCTCGAGGACAGCGGACGAGGTGA	ACATCTGTGACCGACAGCCAGTACGCCCT	2174	
DB	3544	CCTCGCTTTCAGGACTCGGGCTGGAGGTGA	ACATCTGTGACCGACAGTATGCCCT	3603	
QY	2175	GGGCACTATCCAGGCCAGGCCAAGAGCGAG	CGGAGCTGGTGAAACGAGATCATCGA	2234	
DB	3604	GGGCACTATTCAGGCCAGGCCAAGAGCGAG	CGGAGCTGGTGAAACGAGATCATCGA	3663	
QY	2235	GCAGCTGATCAAGAGGAGAGGTGTACTTG	AGCTGGGTGCCCGCCCAACAGGGCATCGG	2294	
DB	3664	GCAGCTGATCAAGAGGAGAGAGGTGTACTT	GGGCTGGGTACCGCCCAACAGGGCATCGG	3723	
QY	2295	CGGCAACGAGCAGATCGACAAGCTGGTG	AGCAAGGGCATCCCGCAAGGTGCTGTCTGTGA	2354	



Db 3896 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACTAA 3955

Qy 1755 CGAGTGAAGCAGCTGACCGAGGCGGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTG 1814

Db 3956 CGAGCTCAAGCAGCTGACCGAGGCGGTGCAGAAGATCACCAACCGAAAGCATCGTGATCTG 4015

Qy 1815 GGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTGGAGACCTGGTGAC 1874

Db 4016 GGGAAAGACTCTTAAGTTCAAGCTGCCCATCCAGAAAGAAACCTGGGAAACCTGGTGAC 4075

Qy 1875 CGACTACTGGCAGGCACTCGATCCCGAGTGGGAGTTCTGTGAACACCCGCCCTGGT 1934

Db 4076 AGAGTATGGCAGGCCACTTGATCTCTAGTGGGAGTTCTGTCAACACCCCTGCCCTGGT 4135

Qy 1935 GAAGCTGTGGTACCAAGCTGGAGAGGAGCCCATCATCGGCCCGAGACCTTCTACGTGA 1994

Db 4136 GAAGCTGTGGTACCAAGCTGGAGAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGA 4195

Qy 1995 CGGCGCCGCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGGGCG 2054

Db 4196 TGGGGCGCTTAACAGGGAGACTAAGCTGGGCAAGCGCGGATACGTCAACCGGGGCGAG 4255

Qy 2055 GCAGAAGATCTGAGCTGACCGAGACCAAGACCAACCAAGAGACCGAGCTGCAGGCGCATCCA 2114

Db 4256 ACAGAAGTTGTACCTCTACTGACACCAACCAAGAGACTGAGCTGCAGGCGCATTTA 4315

Qy 2115 GCTGGCCCTCAGGACAGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGGCCT 2174

Db 4316 CCTCGCTTTCAGGACTCGGCGCTGAGGTGAACATCGTGACAGACTCTCAGTATGCCCT 4375

Qy 2175 GGGCATATCCAGGCCCGACCGCGCAAGAGCGAGCGAGCTGGTGAAACAGATCATCGA 2234

Db 4376 GGGCATATTCAGGCCCGACCGACAGAGTGAGTGCAGCTGGTCAATCAGATCATCGA 4435

Qy 2235 GCAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGGCGCCCGCACAGGCGCATCGG 2294

Db 4436 GCAGCTGATCAAGAGGAGAAAGGTCTATCTGGCTGGGTACCGCGCCACAAAGGCAATGG 4495

Qy 2295 CGGCAACGAGCAGATCGACAGCTGGTGAGCAAGGGGATCGCAGAGGTGCTGTTCCTGGA 2354

Db 4496 CGGCAATGAGCAGTTCGACAGCTGGTCTCGGCTGGCATCAGGAAGGTGCTATTCCTGGA 4555

Qy 2355 CGGCATCGA 2363

Db 4556 TGGCATCGA 4564

RESULT 13

US-09-872-733A-6

; Sequence 6, Application US/09872733A

; Patent No. 6656706

; GENERAL INFORMATION:

; APPLICANT: The Government of the United States of America, as

; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, HIV GAG AND

; FILE OF INVENTION: HIV ENV GENES

; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, HIV GAG & ENV

; CURRENT APPLICATION NUMBER: US/09/872,733A

; CURRENT FILING DATE: 2001-06-01

; PRIOR APPLICATION NUMBER: PCT/US00/34985

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/173,036

; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 8366

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence

; OTHER INFORMATION: of the construct pcwvaggpolENkan containing a CMV

; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin

; OTHER INFORMATION: resistance gene

US-09-872-733A-6

Query Match 63.6%; Score 1566; DB 4; Length 8366;

Best Local Similarity 80.2%; Pred. No. 2e-258;

Matches 1885; Conservative 0; Mismatches 446; Indels 19; Gaps 4;

Qy 33 GGCACACAGCGCCAAACATCTGTATGTCAGGCGCAGCAACTTCAAGGGCCCCCAAGCGCATCAT 92

Db 1879 GACGAACCTCGCGCGACCAATATGATGTCAGAGAGGCAACTTCCGGAACCCACAGCGGAAGATCGT 1938

Qy 93 CAAGTGTCTCAACTGCGCGCAAGGAGGCGCACATCGCCGCAACTGCGCGCGCCCCCGCAA 152

Db 1939 CAAGTGTCTCAATTTGTGCGCAAGGAGGCGCACCGCCAGGAACTGCGCGGCCCCCGGAA 1998

Qy 153 GAAGGGCTGCTGGAAGTGC CGGCAAGAGGGGCCACACAGATGAAGGACTGCACCGAGCGCA 212

Db 1999 GAAGGGCTGTTGGAAATGTTGGAAAGGAGGACACCAATGAAGATTGTACTGAGAGACA 2058

Qy 213 GGCACAACTTCTTCGCGAGGACCTGGCTTCCCCCAGGGCAAGGCGCGGAGTTCCCCAG 272

Db 2059 GGCTAA-TTTTTTTAGGGAAGATCTGGCTTCTCTACAAGGGAAGGCCAGGGAATTTTCTTC 2117

Qy 273 CGAGCAGAACCGCGCCCAACAGCCCCACAGCGCGAGCTGCAGGTGCGCG- ----CGA 326

Db 2118 AGAGCAGACAGAGCCCAACAGCCCCACAGAAAGAGGCTTCAGGTCTTGGGTAGAGACA 2177

Qy 327 CAACCCCGCAGCAGGCGCGCGCGCGCGCGCA-----CCCTGAACTTTCCCCCA 380

Db 2178 CAACTCCCCCTCAGAGCAGGAGCGCGGATAGACAAGGAACTGTATCTTAACTTTCCCTCA 2237

Qy 381 GATCACCTCTGTCAGCGCCCCCTTGGTGAGCATCAAGTGGGCGGCGCAGATCAAGAGGCG 440

Db 2238 GATCACTCTTTGGCAACGACCCCTCGTCAAGTAAGGATCGGGGGGCAACTCAAGGAAGC 2297

Qy 441 CTGCTGGAACACCGCGCGCGCGCGCGCTGCTGGAGGAGATGACCTTGGCCCGCAAGTG 500

Db 2298 GCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAGAAATGATTTGGCAGGAAGATG 2357

Qy 501 GAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGCGCGCGCGCGCAAGCAT 560

Db 2358 GAAACCAAAATGATAGGGGGGATCGGGGGCTTCATCAAGGTGAGGCGAGTACGACCAT 2417

Qy 561 CTGATCGAGATCTGCGCGCAAGAGGCGCATCGGCACCGTGTCTGATCGGCCCGCACCCCGT 620

Db 2418 ACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTCACCTGT 2477

Qy 621 GAACATCATCGCGCGCAACATGCTGACCGAGCTGGGCTGCACCTCTGAACTTTCCCATCAG 680

Db 2478 CAACATAATTGGGAAGAAATCTGTTGACCCAGATCGGCTGCACCTTGAACCTTCCCATCAG 2537

Qy 681 CCCCATCGAGACCGTGCCTGGAAGCTGAAGCGCGCGCATGGAAGCGCGCGCGCGCAAGCA 740

Db 2538 CCTATTGAGACGGTGCCTGGAAGTTGAAGCGGGGATGGAAGCGCGCGCGCGCGCAAGCA 2597

Qy 741 GTGGCCCTTACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGGAGAA 800

Db 2598 ATGGCCATTGACGAAAGAGAGATCAAGGCGCTTAGTCCAAATCTGTACAGAGATGGAGAA 2657

Qy 801 GGAGGCAAGATCACCAAGATCGGCGCGGAGAACCCCTACAAACCCCGCTGTTTCCCAT 860

Db 2658 GGAAGGGAAGATCAGCAAGATCGGGCTTGAGAACCCCTACAAACCTCAGTCTTTCGCAAT 2717

Qy 861 CAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGGAACAAGCG 920

Db 2718 CAAGAAGAAGACAGTACCAAGTGGAGAAAGTGTGTGACTTCAGAGAGCTGAACAAGAG 2777

Qy 921 CACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCCCACCCCGCGCGCTGGAAGAGAA 980

Db 2778 AACTCAGGACTTCTGGGAAGTTTCAAGTGGGCGATCCACATCCCGCTGGTGGTGAAGAGAA 2837

Qy 981 GAAGAGGTGACCGTGTGACAGCTGGGCGAGCGCTACTTACAGGTGCGCTTGGACGAGGA 1040

Db 2838 GAAGTCAGTGACAGTGTGATGTGGGTGATGCTACTTCTCCGTTCCTTGGACGAGGA 2897





1384 TTCAAGTCTGGGTAGACAAACAACTCCCTCAGAAAGAGGACCCGATAGACAAGGAA 1443  
1385 -----CCCTGAACTTCCCTCCAGATCACTCTGTGGCAGCGCCCTCTGTGAGCATCAAGG 418  
1444 CTGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCTGTACAGTAAGGA 1503  
419 TGGGGGCCAGATCAAGGAGCCCTCTGTGGACACCGGGCCGACGACACCGTGTGGAGG 478  
1504 TCGGGGGGCAACTCAAGGAAGCGTGTCTCGATACAGGAGCAGATGATACAGATTAGAAG 1563  
479 AGATGAGCTCCCGGCAAGTGAAGCCCAAGATGATCGGGCGCATCGGGCTTTCATCA 538  
1564 AATGAGTTTCCAGGAAGATGGAACCAAAATGATAGGGGGATCGGGGGCTTTCATCA 1623  
539 AGTGCGCCAGTACGACACAGATCTCTGATTCGAGATCTCGCGCAAGAAGCCATCGGCAACG 598  
1624 AGGTGAGGACGATCGACCCAGATCTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 1683  
599 TGCTGATCGGCCCCACC-----CCGTGAACATCATCGGCCCAACATGCTGACCCAGCTG 654  
1684 TATTAGTAGGACCTTACCTACACCTGTCAACATAATTGGAAGAAATCTGTGACCCAGATC 1743  
655 GGCTGCACCTTGAATCTCCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCC 714  
1744 GGCTGCACCTTGAATCTCCCATCAGCCCTATTGAGACGGTGCCTGTGAAGTTGAAGCCG 1803  
715 GGCAATGAGCGCCCAAGGTGAAGCAGTGGGCCCTGACCGAGGAGAGATCAAGGCCCTG 774  
1804 GGGATGAGCGGCCCAAGGTCAAGCAATGGCCATTGACGAAGAGAGATCAAGGCCCTTA 1863  
775 ACCGCCATCTCGAGGAGATGGAAGAGGAGGCGCAAGATCACCAGATCGGCCCGAGAAC 834  
1864 GTCGAAATCTGTACAGAGATGGAAGAGGAGGAGGAGATCAGCAAGATCGGCCCTGGAAC 1923  
835 CCCTACAAACCCCGCTGTTCGCATCAAGAAAGAGACACCAAGTGGCGCAAGCTG 894  
1924 CCCTACAAACCTCCAGTCTTCGCATCAAGAAAGAGACATACCAAGTGGAGAAAGCTG 1983  
895 GTGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCATC 954  
1984 GTGACTTCAGAGAGTGAACAAGAGAACTCAGGACTTCTGGGAAGTTTCACTGGGCATC 2043  
955 CCCACCCCGCGCTGAAGAAGAGAGAGCGTGACCGTGTGGAGCTGGGCGACGCC 1014  
2044 CCAATCCCGCTGGTGAAGAAGAGAACTCAGTCAAGTGTGGATGTGGGTGATGCC 2103  
1015 TACTTCAGCGTCCCTGGAGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGC 1074  
2104 TACTTCTCCGTTCCCTTGGACGAGACTTCAGGAAGTACACTGCTTCAAGATACCTTAGC 2163  
1075 ATCAACAAACGAGACCCCGCATCCGTTACCAAGTACCAACGCTGCTGCCCGAGGCTGGAAG 1134  
2164 ATCAACAAACGAGACCCAGGATCCGTTACCAAGTACCAAGTGTGCTGCCACAGGATGAG 2223  
1135 GGCAGCCCGAGATCTTCCAGAGCAGATGACCAAGATCTGAGAGCCCTTCCGCGCCCGC 1194  
2224 GGATCACAGCCATCTTTCAAAGCAGATGACCAAGATCTGAGAGCCCTTCCGCAAGCAA 2283  
1195 AACCCTGAGATCTGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGCGACTGGAG 1248  
2284 AACCCTGAGATCTGTATCTATCACTACATGAGCAACCTCTACGTGAGGAAGTGAACCTGGAG 2343  
1249 ATC-GGGCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTCTGCTGGCGCTGGGGCTT 1307  
2344 ATCGGGGCGACACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGTTGGGACT 2403  
1308 CACCACCCCGAGCAAGAGAGACCAAGAGGAGCCCTTCTCTGTGATGAGGCTACGAGCT 1367  
2404 GACCACACCAAGCAAGAGACCAAGAGGAACTCTCCCTTCTGTGGATGGGCTACGAAC 2463  
1368 GCACCCCGCAAGTGGAGCCCTGAGCCCATCGAGCTGCCGAGAGGAGGAGCTGGACCGT 1427  
2464 GCATCTGCAAGTGGAGCAGTGGACGCCCATCTGCTGCTGAGAGGAGACAGCTGGACTGT 2523

Qy 1428 GAACGACATCCAGAAAGCTGTGTGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCCGGAT 1487  
Db 2524 GAAACGACATCAGAAAGCTGTGTGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCAGGAT 2583  
Qy 1488 CAAGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCAAGGCCCTGACCCGACATCTGTGCC 1547  
Db 2584 CAAAGTTAGGAGCTGTGCAAGCTGTGCAAGCTGTGCAAGGCCCTGACCCGACATCTGTGCC 2643  
Qy 1548 CCTGACGAGGAGCCGAGCTGTGAGCTGGCCGAGAACCGCGAGATCTTGCCTGCGAGCCCGT 1607  
Db 2644 ACTGACAGAGGAGCAGAGCTAGAACTTGGCAGAGAACCGAGAGATCTTGAAGGAGCCAGT 2703  
Qy 1608 GCACGGCTGTATCTACGACCCCAAGCAAGCACTTGGTGGCCGAGATCTCAAGAGCAGGCGCA 1667  
Db 2704 ACATGAGTGTACTTACGACCCCAAGCAAGCACTTGTGCGCAGAGATCTCAAGAGCAGGCGCA 2763  
Qy 1668 CGACAGTGGACCTTACAGATCTTACCAAGGAGCCCTTCAAGAACTTGAAGCAGGCAAGTA 1727  
Db 2764 AGGCAATGGACCTTACCAAAATCTTACCAAGGAGCCCTTCAAGAACTTGAAGCAGGCAAGTA 2823  
Qy 1728 CGCAGATGCGCACCCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAGAA 1787  
Db 2824 CGCAAGATGAGGGTGCCTCCACACCAACGATGTGAAGCAGCTGACAGGCGAGTGCAGAA 2883  
Qy 1788 GATCGCATGGAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCCGCTTGCCTCATCCA 1847  
Db 2884 GATCACACAGAGAGCATCTGTGATCTGGGCGAAGACTCCCAAGTTTCAAGCTGCCCATACA 2943  
Qy 1848 GAGGAGACCTGGGAGACCTTGTGGACCGACTACTGCGCAGCCACCTGGATTCCTCGAGTG 1907  
Db 2944 GAGGAGACATGGGAGACATGTGGACCGAGTACTGGCAAGCCACCTGGATTCCTCGAGTG 3003  
Qy 1908 GGAGTTCTGTAACACCCCTTGTGTGAAGCTGTGTGTACAGCTGGAAGAGGAGCCCAT 1967  
Db 3004 GGAGTTCTGTAACACCCCTTGTGTGAAGCTGTGTGTATCAGCTGGAAGAGGAGCCAT 3063  
Qy 1968 CATCGCGCGCAGACCTTCTACGTGTGACGCGCGCCCAACCGCGAGACCAAGATCGGCAA 2027  
Db 3064 CGTGGAGCAGAGACCTTCTACGTGTGATGGGCGAGCCAAACAGGGAGACCAAGCTGGGCAA 3123  
Qy 2028 GCGCGGCTACGTGACCGACCGGGCGCGGAGAGATGTGTGAGCTGTGAGCTGACCGAGACCA 2087  
Db 3124 GCGAGGCTACGTGACCAACCGGAGACGACAAAGTGTGTGAGCTGTGACCTGTGACACCA 3183  
Qy 2088 CCAGAACCGCAGCTGCAAGCTGAGCTGGCCCTGCGAGCAGCGGCGAGCGAGCTGAGTCAA 2147  
Db 3184 CCAGAACCTGTGAGCTGCAAGCTTACCTAGCTTCTGCAAGACAGCGGACTGGAAGTCAA 3243  
Qy 2148 CATCGTACCGCAGCAGCTACGCTTGGGCTCATCTCAGGCCCGAGCCCGCAAGAGCGA 2207  
Db 3244 CATCGTACAGACTCAAGTACG-CATGGGCTCATCTCAAGCACAACAGACCAATCCGA 3302  
Qy 2208 GAGGAGCTGTGACAGCATCTGAGCAGCTGATCAAGAGGAGAGGAGTGTACTCTGAG 2267  
Db 3303 GTCAGAGCTGTGAGCAGATCTGAGCAGCTGATCAAGAGGAGGAGAGTGTACTCTGGC 3362  
Qy 2268 CTGGGTCCCGCCCAAGGCGCATCTGGCGGCAAGCAGCAGATCGACAGCTGTGTGAGCAA 2327  
Db 3363 ATGGGTACGAGCACAAGAGGATTTGAGGAGGATGAACAGTGTAGTAAATTAGTCAGTGC 3422  
Qy 2328 GGGCATCCGCAAGTGTGTCTTCTGAGCGGATCGAT 2364  
Db 3423 TGGGATCCGGAAGGTGTCTTCTCTGGACGGATCGAT 3459

RESULT 15  
US-09-952-060-1  
; Sequence 1, Application US/09952060  
; Patent No. 6733993  
; GENERAL INFORMATION:  
; APPLICANT: Emini, Emilio A.  
; APPLICANT: Youil, Rima

<p> ; APPLICANT: Bett, Andrew J.  ; APPLICANT: Chen, Ling  ; APPLICANT: Kaslow, David C.  ; APPLICANT: Shiver, John W.  ; APPLICANT: Toner, Timothy J.  ; APPLICANT: Casimiro, Danilo R.  ; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  ; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  ; TITLE OF INVENTION: MODIFICATIONS  ; FILE REFERENCE: 20747Y  ; CURRENT APPLICATION NUMBER: US/09/952,060  ; CURRENT FILING DATE: 2001-09-14  ; PRIOR APPLICATION NUMBER: PCT/US01/28861  ; PRIOR FILING DATE: 2001-09-14  ; PRIOR FILING DATE: 2001-09-07  ; PRIOR APPLICATION NUMBER: 60/317,814  ; PRIOR FILING DATE: 2001-09-07  ; PRIOR APPLICATION NUMBER: 60/279,056  ; PRIOR FILING DATE: 2001-03-27  ; PRIOR APPLICATION NUMBER: 60/233,180  ; PRIOR FILING DATE: 2000-09-15  ; NUMBER OF SEQ ID NOS: 38  ; SOFTWARE: FastSeq for Windows Version 4.0  ; SEQ ID NO 1  ; LENGTH: 2577  ; TYPE: DNA  ; ORGANISM: Artificial Sequence  ; FEATURE:  ; OTHER INFORMATION: Codon optimized DNA encoding modified wt pol  US-09-952-060-1 </p>		<p> Query Match 53.2%; Score 1309.6; DB 4; Length 2577;  Best Local Similarity 86.2%; Pred. No. 9.6e-215;  Matches 1463; Conservative 0; Mismatches 229; Indels 6; Gaps 1; </p>	
Qy	672	CCCCATCAGCCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGATGAGACCGGCCCAA	731
Db	15	CCCCATCCTCCCCATCGAGACTGTCCCTGTGAAGCTGAAGCCCTGGCATGGATGGCCCAA	74
Qy	732	GGTGAAGCAGTGGCCCTGACCGGAGGAGATCAAGGCCCTGACCGCCATCTGCGAGGA	791
Db	75	GGTGAAGCAGTGGCCCTGACCGGAGGAGATCAAGGCCCTGGTGAATCTGCACTGA	134
Qy	792	GATGGAGAGAGGGGCAAGATCAACAGATCGGCCCGGAGAACCCCTTACAAACACCCCGCT	851
Db	135	GATGGAGAGAGGGGCAAAATCTCAAGATGGCCCCCGAGAACCCCTTACAAACACCCCTGT	194
Qy	852	GTTCGCCATCAAGAAGAGACACCAAGTGGCGCAAGCTGGTGAATCTTCGCGAGCT	911
Db	195	GTTCGCCATCAAGAAGAGAGCTCCACCAAGTGGAGAGAGCTGGTGAATCTTCAGGGAGCT	254
Qy	912	GAACAGCGCACCCAGACCTCTGGGAGGTGCAGCTGGGATCCCGCCCGCGGCGCT	971
Db	255	GAACAGAGAGACCCAGACCTCTGGGAGGTGCAGCTGGGATCCCGCCCGCGGCGCT	314
Qy	972	GAAGAAGAGAGAGCTGACCGCTGCTGGAGCGTGGGCGAGCGCTTCTTCAGCGTGGCCCT	1031
Db	315	GAAGAAGAGAGAGCTGCTGACTGTGCTGGATGGGGATGCTTCTCTGTCGCCCT	374
Qy	1032	GGACGAGGACTTCGCGAAGTACACCGCTTTCACATCCCGAGATCAACAAACGAGACCCC	1091
Db	375	GGATGAGGACTTCAGGAAGTACACTGCTCTTCAACATCCCTCCATCAACAAATGAGACCCC	434
Qy	1092	CGGCATCCGCTACAGTACAACTGTGCTCCCGAGGCTGAAGGCGAGCCCGAGCATCTT	1151
Db	435	TGGCATCAGGTACAGTACAACTGTGCTCCCGAGGCTGAAGGCGAGCCCGAGCATCTT	494
Qy	1152	CCAGAGCAGCATCAACAAAGTCTTGGAGCCCTTTCGCGCGCCGCAACCCCGAGATCTGTAT	1211
Db	495	CCAGTCTCATGACCAAGATCTGGAGCCCTTTCAGGAAGCAGNACCTTGACATTTGTAT	554
Qy	1212	CTACCA-----GGCCCCCTGTGTACGTGGGAGCGACTCGAGATCGGCCAGCACCGCGC	1265
Db	555	CTACCAAGTACATGGATGACCTGTATGTGGGCTGTGACCTGGAGATTTGGGAGCACAGGAC	614

Qy	1266	CAAGATCGAGGAGCTGCGAAGCACCTGCTGGCTGGGGCTTCCACCACCCCCGACAAGAA	1325
Db	615	CAAGATTCGAGGAGCTGAGCGACACCTGCTGAGTGGGGCTGACCAACCCCTGACAAGAA	674
Qy	1326	GCACCAAGAGAGCCCCCTTCTGTGGATGGGCTACGAGCTGCAACCCCGACAAGTGGAC	1385
Db	675	GCACCAAGAGAGCCCCCTTCTGTGGATGGGCTATGAGCTGCAACCCCGACAAGTGGAC	734
Qy	1386	CGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGAACCGTGAACGACATCCAGAGCT	1445
Db	735	TGTGCAGCCCATTTGTCTGCTGAGAAAGGACTCTCTGGAATGAAATGACATCCAGAGCT	794
Qy	1446	GGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTG	1505
Db	795	GGTGGGCAAGCTGAACTGGGCCCTCCCAATCTACCTTGGCATCAAGGTGAGCGAGCTGTG	854
Qy	1506	CAAGCTGTGCGCGGCCCAAGGCCCTGACCGCATCTGTCGCCCTTACCGAGAGAGCCGA	1565
Db	855	CAAGCTGTGAGGGGCAACCAAGGCCCTGACTGAGGTGATCCCCCTGACTGAGGAGGCTGA	914
Qy	1566	GCTGGAGCTGGCCGAGAACCGCGAGATCTGCGGAGCCCGTGCAAGCGGCTGTACTACGA	1625
Db	915	GCTGGAGCTGGCTGAGAACAGGGAGATCTTGAAGGAGCCCTGTGCATGGGGTGTACTATGA	974
Qy	1626	CCCACAGAGGACCTGGTGGCCGAGATCCAGAGAGAGGGCCAGCACAGTGGACCTTACCA	1685
Db	975	CCCTCTCAGGACCTGATTTGCTGAGATCCAGAGAGAGGGCCAGGGCCAGTGGACCTACCA	1034
Qy	1686	GATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGC	1745
Db	1035	AATCTACAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCCAGGATGAGGGGGG	1094
Qy	1746	CCACACAGGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCCATGAGAGCAT	1805
Db	1095	CCACACCAATGATGTGAAGCAGCTGACTGAGGCTGTGCAAGAGATCACCACCTGAGTCCAT	1154
Qy	1806	CGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTTGGGAGAC	1865
Db	1155	TGTGATCTGGGGCAAGACCCCAAGTTCAAGTGTGCCCATCCAGAGAGACCTTGGGAGAC	1214
Qy	1866	CTGGTGGACCGACTACTTGGCAGGCCACCTGGATTCCTGGAGTGGGAGTTCGTGAACACCC	1925
Db	1215	CTGGTGGACTGAGTACTTGGCAGGCCACCTGGATTCCTGGAGTGGGAGTTCGTGAACACCC	1274
Qy	1926	CCCCCTGGTGAAGCTGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAGACCTT	1985
Db	1275	CCCCCTGGTGAAGCTGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAGACCTT	1334
Qy	1986	CTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGCTTACGTCACCGA	2045
Db	1335	CTATGTGGATGGGCTGCCCAACAGGGAGACCAAGCTGGGCAAGGCTTGTGACCAA	1394
Qy	2046	CGGGGCCCGCAGAAAGATCGTGAAGCTGACCGAGACCAACCAACAGAAAGACCGAGTGA	2105
Db	1395	CAGGGCGAGCAGAAAGTGGTGACCTGACTGACACCAACCAACAGAAAGACTGAGTCCA	1454
Qy	2106	GGCCATTCAGCTGGCCCTGCAAGACAGCGCAGAGGTGAACATCGTGACCGCAGACCA	2165
Db	1455	GGCCATTCATCTGGCCCTTCAGGACTCTTGGCTTGGAGGTGAACATTTGTGACTGACTCCA	1514
Qy	2166	GTACGCCCTGGGATCATCTCAGGCCCGAGCCCAAGAGAGCGAGAGGAGTGGTGAACCA	2225
Db	1515	GTATGCCCTGGGATCATCTCAGGCCCGAGCTGATCAGTCTGAGTCTGAGCTGGTGAACCA	1574
Qy	2226	GATCATCGACAGCTGATCAAGAGAGAGAGGTGTAACCTGAGCTGGGTGCCCCCACA	2285
Db	1575	GATCATCGACAGCTGATCAAGAGAGAGAGGTGTACTCTGGCTGGGTGCTGCCCAACA	1634
Qy	2286	GGGCATCGGGCGCAACGAGCAGATCCAGAGCTGTGAGCAAGGGGATCCGCAAGGTGCT	2345
Db	1635	GGGCATTTGGGGGCAATGAGCAGGTGCAAGCTGGTGTCTGTGGCATCAGGAAGGTGCT	1694

Qy 2346 GTTCCTGGACGCGATCGA 2363  
|||||||  
Db 1695 GTTCCTGGATGCGATTGA 1712

Search completed: June 2, 2005, 03:13:05  
Job time : 401 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 09:47:50 ; Search time 388.052 Seconds  
(without alignments)

10360.282 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457

Sequence: 1 gtcagccaccatggcga.....gggtagcaccggtgaattc 2457

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2040	83.0	2300	4	US-09-475-515-83
2	2039.2	82.2	2306	4	US-09-475-515-82
3	1998.4	81.3	2312	4	US-09-475-515-84
4	1915.8	78.0	4319	4	US-09-475-515-6
5	1872.8	76.2	2299	4	US-09-475-515-81
6	1852	75.4	2305	4	US-09-475-515-80
7	1626.6	66.2	4307	4	US-09-552-950-2
8	1607.4	65.4	4307	4	US-09-936-572-2
9	1599.4	65.1	4327	4	US-09-936-572-14
10	1599.4	65.1	4353	4	US-09-936-572-13
11	1599.4	65.1	4642	4	US-09-936-572-12
12	1599.4	65.1	9772	4	US-09-552-950-5
13	1539.2	62.6	8366	4	US-09-872-733A-6
14	1503.8	61.2	4338	4	US-09-872-733A-1
15	1282.8	52.2	2577	4	US-09-952-060-1
16	1282.8	52.2	2650	4	US-09-952-060-5
17	1280.2	52.1	4053	4	US-09-952-060-34
18	1278	52.0	2577	4	US-09-952-060-3
19	1278	52.0	2650	4	US-09-952-060-7
20	1278	52.0	38519	4	US-09-952-060-28
21	1178.4	48.0	9010	4	US-09-184-418C-8
22	1150.6	46.8	9913	4	US-09-827-688-11
23	1147.4	46.7	8972	4	US-09-184-418C-9
24	1140.2	46.4	8959	4	US-09-184-418C-11
25	1137	46.3	2467	4	US-09-872-733A-3
26	1116.8	45.5	8992	4	US-09-184-418C-4
27	1081.6	44.0	2601	3	US-09-117-217-7

28 1081.6 44.0 2601 3 US-09-117-217-9 Sequence 9, Appli  
29 1081.6 44.0 2601 3 US-09-117-217-11 Sequence 11, Appli  
30 1081.6 44.0 2601 3 US-09-117-217-13 Sequence 13, Appli  
31 1081.6 44.0 2601 4 US-09-735-487-7 Sequence 7, Appli  
32 1081.6 44.0 2601 4 US-09-735-487-9 Sequence 9, Appli  
33 1081.6 44.0 2601 4 US-09-735-487-11 Sequence 11, Appli  
34 1081.6 44.0 2601 4 US-09-735-487-13 Sequence 13, Appli  
35 1081.6 44.0 4307 4 US-09-552-950-1 Sequence 1, Appli  
36 1081.6 44.0 4307 4 US-09-936-572-1 Sequence 1, Appli  
37 1081.6 44.0 9719 4 US-09-700-304-1 Sequence 1, Appli  
38 1080 44.0 9050 4 US-09-184-418C-7 Sequence 7, Appli  
39 1075.2 43.8 7399 2 US-08-418-848A-9 Sequence 9, Appli  
40 1075.2 43.8 9709 2 US-08-188-583-5 Sequence 5, Appli  
41 1075.2 43.8 9709 3 US-08-388-353-1 Sequence 1, Appli  
42 1075.2 43.8 9709 3 US-08-488-551B-1 Sequence 1, Appli  
43 1075.2 43.8 9709 3 US-09-309-572-15 Sequence 15, Appli  
44 1075.2 43.8 9709 4 US-09-718-096-15 Sequence 15, Appli  
45 1075.2 43.8 12473 4 US-09-318-138-13 Sequence 13, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-475-515-83

; Sequence 83, Application US/09475515A

; Patent No. 6602705

; GENERAL INFORMATION:

; APPLICANT: BARNETT, Susan

; APPLICANT: ZUR MEGEDE, Jan

; APPLICANT: SRIVASTAVA, Indresh

; APPLICANT: LIAN, Ying

; APPLICANT: HARTOG, Karin

; APPLICANT: LIU, Hong

; APPLICANT: GREER, Catherine

; APPLICANT: SELBY, Mark

; APPLICANT: WALKER, Christopher

; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION

; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES

; FILE REFERENCE: 1621.002

; CURRENT APPLICATION NUMBER: US/09/475,515A

; CURRENT FILING DATE: 1999-12-30

; NUMBER OF SEQ ID NOS: 90

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 83

; LENGTH: 2300

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:

; OTHER INFORMATION: FS(-).protmod.RTopt.YMM

US-09-475-515-83

Query Match 83.0%; Score 2040; DB 4; Length 2300;

Best Local Similarity 93.6%; Pred. No. 0;

Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

Qy 170 GCGCAGGAGGGCCACCATGAGGACTGCAGCGCCGAGCCAGCAACTTCTTCGCG 229  
Db 1 GCGCCCGGAGGACACCAATGAAGATTGCACTGAGAGACAGCTAATTTCTTCGCG 60  
Qy 230 AGGACCTGGGCTTCCCCCAGGCGCAAGCCCGCGAGTTCCCCCAGGAGCAGAACCGCGCA 289  
Db 61 AGGACCTGGGCTTCTCTGCGGCAAGCCCGCGAGTTTCAGCAGGAGCAGACCCGCGCA 120  
Qy 290 ACAGCCCCACAGCCGCGAGCTGCAGTGGCGCGCG-----ACAACCCCGCAGCGAGG 343  
Db 121 ACAGCCCCACCGCGCGAGCTGCAGTGTGGGCGCGGAGAACACAGCTGAGCGAGG 180  
Qy 344 CCGCGCGCGAGCGCCAGGCGACCCCTG-----ACTTCCCCCAGATCACCCTGTGGCAGC 397  
Db 181 CCGCGCGCGAGCGCCAGGCGACCGGTGAGCTTCAACTTCCCCCAGATCACCCTGTGGCAGC 240



QY 398 GCCCTCGTGGAGCATCAAGTGGCGGCAGATCAAGAGAGCCCTGTGTGACACCGGC 457  
 Db 241 GCCCTCGTGGAGCATCAAGTGGCGGCAGATCAAGAGAGCCCTGTGTGACACCGGC 300  
 QY 458 CGGACGACCCGTGTGGAGGAGATGAGCTTCCCGGCAAGTGGAGGCCCAAGATGATCG 517  
 Db 301 CGGACGACCCGTGTGGAGGAGATGAGCTTCCCGGCAAGTGGAGGCCCAAGATGATCG 360  
 QY 518 GGGGATCGGCGGCTTCAAGAGTGGCGCAGTACGACAGATCTGTGATCGAGATCTGG 577  
 Db 361 GGGGATCGGCGGCTTCAAGAGTGGCGCAGTACGACAGATCTGTGATCGAGATCTGG 420  
 QY 578 GCAAGAGCCATCGGCGGCTGATCGGCGCCACCCCGTGAACATCATCGGCGCA 637  
 Db 421 GCAAGAGCCATCGGCGGCTGATCGGCGCCACCCCGTGAACATCATCGGCGCA 480  
 QY 638 ACATGTCACCCAGCTGGGCTGCACCCCTGAACTTCCCATCAGCCCATCGAGCCGTGC 697  
 Db 481 ACCTGTCACCCAGATCGGCTGCACCCCTGAACTTCCCATCAGCCCATCGAGCCGTGC 540  
 QY 698 CGGTGAAGTGAAGCCCGGATGAGCGGCCCAAGGTGAAGCACTGTGAGCCCTGACCGAG 757  
 Db 541 CGGTGAAGTGAAGCCCGGATGAGCGGCCCAAGGTGAAGCACTGTGAGCCCTGACCGAG 600  
 QY 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGAGATGAGAGAGGAGGAGGACAGCA 817  
 Db 601 AGAAGATCAAGGCGCTGAGCGCATCTGCGAGAGATGAGAGAGGAGGAGGAGGACAGCA 660  
 QY 818 AGATCGGCGCCGAGAACCCCTACAACACCCCGCTGTTCGCCATCAAGAGAGGAGCAGCA 877  
 Db 661 AGATCGGCGCCGAGAACCCCTACAACACCCCGCTGTTCGCCATCAAGAGAGGAGCAGCA 720  
 QY 878 CGAAGTGGCGAAGCTGTGTGATCTTCGGGAGCTGAAACAGCGCACCCGAGCTTCGGG 937  
 Db 721 CGAAGTGGCGAAGCTGTGTGATCTTCGGGAGCTGAAACAGCGCACCCGAGCTTCGGG 780  
 QY 938 AGGTGAGCTGGGCTGAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 997  
 Db 781 AGGTGAGCTGGGCTGAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
 QY 998 TGGAGCTGGGCGACCTACTTCAAGCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAG 1057  
 Db 841 TGGAGCTGGGCGACCTACTTCAAGCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAG 900  
 QY 1058 CTTTCAACATCCCGAGATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117  
 Db 901 CTTTCAACATCCCGAGATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 QY 1118 TGCCCGAGGCTGGAAGGCGAGCCAGCATCTTCAGAGAGAGAGAGAGAGAGAGAG 1177  
 Db 961 TGCCCGAGGCTGGAAGGCGAGCCAGCATCTTCAGAGAGAGAGAGAGAGAGAGAG 1020  
 QY 1178 AGCCCTTCGCGCCGCGCAACCCCGAGATCGTGTATCTACAGGAGAGAGAGAGAGAG 1237  
 Db 1021 AGCCCTTCGCGCCGCGCAACCCCGAGATCGTGTATCTACAGGAGAGAGAGAGAGAG 1080  
 QY 1238 GGGAGCTGAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297  
 Db 1081 GGGAGCTGAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 QY 1298 GGTGGGCTTACACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357  
 Db 1141 GGTGGGCTTACACCCCGAG 1200  
 QY 1358 AGCTGCACCCCGAG 1417  
 Db 1201 AGCTGCACCCCGAG 1260  
 QY 1418 CGGTGAACGACATCGAG 1477  
 Db 1261 CGGTGAACGACATCGAG 1320  
 QY 1478 GCATCAAGTGTGGCGGAGCTGTGCAAGTGTGCGGCGGCGGAGGCGGCTGACCGCATCG 1537

Db 1321 GCATCAAGTGTGAGCAGCTGTGCAAGCTGTGCGCGGCACCAAGGCGCTGACCGAGTGA 1380  
 QY 1538 TGCCCTTCACCGAGAGAGCCGAGCTGTGAGCTGTGCGCGAGAAACCGCGAGATCTTCTGCGGAGC 1597  
 Db 1381 TCCCTTCACCGAGAGAGCCGAGCTGTGAGCTGTGCGCGAGAAACCGCGAGATCTTCTGAAGGAGC 1440  
 QY 1598 CGGTGACGCGGCTGTACTACGACCCACGACAGGAGCTGTGCGCGAGATCTTCAAGAGAGCAGG 1657  
 Db 1441 CGGTGACGAGGTGTACTACGACCCACGACAGGAGCTGTGCGCGAGATCTTCAAGAGAGCAGG 1500  
 QY 1658 GCCACGACAGTGGAGCTTACAGATCTTACAGGAGCCCTTCAAGAGAACCTTGAAGACCGGCA 1717  
 Db 1501 GCCAGGCGAGTGGAGCTTACAGATCTTACAGGAGCCCTTCAAGAGAACCTTGAAGACCGGCA 1560  
 QY 1718 AGTACGCAAGATGCGCACCGGCCACACGACGAGAGAGCTGAGAGAGAGAGAGAGAGAGAG 1777  
 Db 1561 AGTACGCGATGCGCGCGGCCACACGACGAGAGAGCTGTGCGCGAGATCTTCAAGAGAGCAGG 1620  
 QY 1778 AGAGATCGCATGGAGAGAGCTGTGATCTGGGCAAGACCCCAAGTTCGCCCTGCCCA 1837  
 Db 1621 AGAGGTGAGCACCGAGAGAGCTGTGATCTGGGCAAGATCTTCAAGAGAACCTTCAAGAGTGC 1680  
 QY 1838 TCCAGAGAGAGACCTTGGGAGAGCTGTGAGAGCCGATCTGCGAGGCGCCACCTGGATCCCCG 1897  
 Db 1681 TCCAGAGAGAGACCTTGGGAGGCTGTGATGAGTGTGCGAGGCGCCACCTGGATCCCCG 1740  
 QY 1898 AGTGGAGATCTGTGAACACCCCGCTGTGAGAGAGCTGTGAGTGTGAGTGTGAGAGAGAGAG 1957  
 Db 1741 AGTGGAGATCTGTGAACACCCCGCTGTGAGAGAGCTGTGAGTGTGAGTGTGAGAGAGAGAG 1800  
 QY 1958 CCATCATCGGCGCGAGAGCTTCTACGTGGAGCGCGCGCCCAACCGCGAGAGCAAGATCG 2017  
 Db 1801 CCATCGTGGGCGCGAGAGCTTCTACGTGGAGCGCGCGCCCAACCGCGAGAGCAAGATCG 1860  
 QY 2018 GCAAGGCGGCTTACGTGACCGACCGGCGCGCGAGAGATCTGAGAGCTGTGAGAGAGAGAG 2077  
 Db 1861 GCAAGGCGGCTTACGTGACCGAGCCGCGCGCGAGAGCTGTGAGAGCTGTGAGAGAGAGAG 1920  
 QY 2078 CCAACGAGAGAGAGCTGTGAG 2137  
 Db 1921 CCAACGAGAGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
 QY 2138 TGAACATCGTGGAG 2197  
 Db 1981 TGAACATCGTGGAG 2040  
 QY 2198 GCGAGAGAGAGCTGTGAG 2257  
 Db 2041 GCGAGAGAGAGCTGTGAG 2100  
 QY 2258 TGAGCTGGGTGCGCGCCCAAGGAG 2317  
 Db 2101 TGCGCTGGGTGCGCGCCCAAGGAG 2160  
 QY 2318 GCAAGGAG 2377  
 Db 2161 GCGCGGAG 2220  
 QY 2378 AGTATCGAG 2437  
 Db 2221 AGTATCGAG 2280  
 QY 2438 GGGCTAGCAGCGGTGAATTC 2457  
 Db 2281 GGGCTAGCAGCGGTGAATTC 2300

RESULT 2  
 US-09-475-515-82  
 ; Sequence 82, Application US/09475515A  
 ; Patent No. 6602705  
 ; GENERAL INFORMATION:

```
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475.515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 2306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YM
US-09-475-515-82

Query Match      82.2%; Score 2019.2; DB 4; Length 2306;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

Qy 170 GCGCAAGAGGGCCACAGATGAGTAAAGACTGACCGAGCGGCAGGCCCAATCTTCTCCGCG 229
Db 1 GCGGCGCGGAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAAATTTCTTCGCG 60

Qy 230 AGGACCTGGCTTCCCGAGGGCAAGCGCCGCGAGTTCCTCCAGCGAGCAAGCCGCCA 289
Db 61 AGGACCTGGCTTCTTGAGGGCAAGCGCCGCGAGTTTCAGCAGGAGCAGACCCGCGCA 120

Qy 290 ACAGCCCCCAGCGCGAGCTGCGAGGTGCGCGCG-----ACAAACCCCGCAGCGAGG 343
Db 121 ACAGCCCCCAGCGCGAGCTGCGAGGTGCGCGCGCGAGAAACAAGCTGAGCGAGG 180

Qy 344 CCGCGCGCGAGCGCGAGGCAACCTG-----AACTTCCCGCAGATACCTCTGGGAGC 397
Db 181 CCGCGCGCGAGCGCGAGGCAACCTGAGCTTCAACTTCCCGCAGATACCTCTGGGAGC 240

Qy 398 GCGGCGGCTGAGCATCAAGTGGCGCGCGAGATCAAGGAGCGCTGCTGAGCAACCGCG 457
Db 241 GCGGCGGCTGAGCATCAAGTGGCGCGCGAGCTCAAGGAGCGCTGCTGAGCAACCGCG 300

Qy 458 CCGAGCACACCGTCTGAGGAGATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGAGCACACCGTCTGAGGAGATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCG 360

Qy 518 GCGGCGTGGCGGCTTCAATCAAGTGGCGCGCGAGTACGACAGATCCTGATCGAGATCG 577
Db 361 GCGGCGTGGCGGCTTCAATCAAGTGGCGCGCGAGTACGACAGATCCTGATCGAGATCG 420

Qy 578 GCAAGAGGCCATCGGCACCGTGTGATCGCGGCCCAACCCCGTGAACATCATCGCGCGCA 637
Db 421 GCAAGAGGCCATCGGCACCGTGTGATCGCGGCCCAACCCCGTGAACATCATCGCGCGCA 480

Qy 638 ACATGCTGACCCAGCTGGGCTGCAACCTTGAACCTTCCCATCAGCCCATCGAGACCGTGC 697
Db 481 ACCTGCTGACCCAGATCGGCTGCAACCTTGAACCTTCCCATCAGCCCATCGAGACCGTGC 540

Qy 698 CCGTGAAGCTGAAGCCCGGCAATGAGCGGCCCAAGTGAAGCAGTGGCCCTGACCGAGG 757
Db 541 CCGTGAAGCTGAAGCGGGGATGAGCGGCCCAAGTGAAGCAGTGGCCCTGACCGAGG 600

Qy 758 AGAAGATCAAGGCTTGAACCCCATCTGCGAGGAGATGAGAGAGGGGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCTTGTGGAGATCTGCAACCGAGATGAGAGAGGGGCAAGATCACCA 660

818 AGATCGCCCGCGAGAAACCCCTTACAAACCCCGCTGTTCGCCATCAAGAAGAGACAGCA 877
661 AGATCGCCCGCGAGAAACCCCTTACAAACCCCGCTGTTCGCCATCAAGAAGAGACAGCA 720

878 CCAAGTGGCGCGAGCTGGTGGACTTCCGCGAGCTGAACAGCGCACCCAGACTTCTGGG 937
721 CCAAGTGGCGCGAGCTGGTGGACTTCCGCGAGCTGAACAGCGCACCCAGACTTCTGGG 780

938 AGGTGACAGCTGGGCGATCCCCCACCACCCCGCGCTTGAAGAAGAAAGAGCGTACCGTGC 997
781 AGGTGACAGCTGGGCGATCCCCCACCACCCCGCGCTTGAAGAAGAAAGAGCGTACCGTGC 840

998 TGGACGTGGGCGAGCGCTTACTTACAGCTGCGCCCTGGAAGAGACTTCCGAAGTACACG 1057
841 TGGACGTGGGCGAGCGCTTACTTACAGCTGCGCCCTGGAAGAGACTTCCGAAGTACACG 900

1058 CTTTACCATCCCCAGCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
901 CTTTACCATCCCCAGCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

1118 TGCCCCAGGGCTGAAGGGGAGCGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1177
961 TGCCCCAGGGCTGAAGGGGAGCGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1020

1178 AGCCCTTCCGCGCGCCCAACCCCGAGATCTGTGATCTACAGGCGCCCTGTGTACGAGCA 1237
1021 AGCCCTTCCGCGCGAGAGAACCCCGAGATCTGTGATCTACAGGCGCCCTGTGTACGAGCA 1080

1238 GCGACCTGGAGATCGGCGAGCAGCGCCAGATGAGAGAGCTTGGCAGAGCAGCTGCTGC 1297
1081 GCGACCTGGAGATCGGCGAGCAGCGCCAGATGAGAGAGCTTGGCAGAGCAGCTGCTGC 1140

1298 GCTGGGCGCTTACCAACCCCGAGAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1355
1141 GCTGGGCGCTTACCAACCCCGAGAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

1356 ----CGAGCTGCAACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1411
1201 GCTACGAGCTGCAACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

1412 GCTGGACCGGTGAACGATCCAGAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCT 1471
1261 GCTGGACCGGTGAACGATCCAGAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCT 1320

1472 ACCCGCGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGGCGCGCAAGAGAGAGAGAGAG 1531
1321 AGCGCGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGGCGCGCAAGAGAGAGAGAGAG 1380

1532 ACATCGTGGCGCTGACCGAGAGAGCGAGCTGGAGCTGGCGCGAGAAACCGCGAGATCTCTGC 1591
1381 AGGTGATCCCCCTGACCGAGAGAGCGCGAGCTGGAGCTGGCGCGAGAAACCGCGAGATCTCTGA 1440

1592 GCGAGCCCGTGCACGCGGTGTATCTACGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1651
1441 AGGAGCCCGTGCACGAGGTGTATCTACGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500

1652 AGCAGGCGCGAGCAGCAGTGGAGCTTACAGAGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAG 1711
1501 AGCAGGCGCGAGGCGAGTGGAGCTTACAGAGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAG 1560

1712 CCGGCAAGTACGCGCAAGATGCGCACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1771
1561 CCGGCAAGTACGCGCGAGTGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620

1772 CCGTGAAGAGATCGCCATGAGAGAGATCTGTGATCTGGGCGAGAGAGAGAGAGAGAGAGAGAG 1831
1621 CCGTGAAGAGATGAGCAGCAGAGAGATCTGTGATCTGGGCGAGAGAGAGAGAGAGAGAGAGAG 1680

1832 TGCCCATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
1681 TGCCCATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740

1892 TCCCGGAGTGGAGTTCGTGAAACACCCCGCGCTGTGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1951
```

```

Db 1741 TCCCGAGTGGAGTTCGTGAACACCCCCCTGGTGAAGCTGTGGTACCACTGGAGA 1800
QY 1952 AGGAGCCATCATCGGCGCGAGAGCTTCTAGTGAACGGCGCCGCAACCGCGAGACA 2011
Db 1801 AGGAGCCATCATCGGCGCGAGAGCTTCTAGTGAACGGCGCCGCAACCGCGAGACA 1860
QY 2012 AGATCGGCAAGCGCGCTACGTGACCGACCGGGCGCGCAGAGATCTGTAGCCTGACCG 2071
Db 1861 AGCTGGGCAAGCGCGCTACGTGACCGACCGGGCGCGCAGAGATCTGTAGCCTGACCG 1920
QY 2072 AGACCAACCAACAGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2131
Db 1921 ACACCAACCAACAGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCC 1980
QY 2132 GCGAGGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCATCATCCAGGCCACGCCG 2191
Db 1981 TGGAGGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCATCATCCAGGCCACGCCG 2040
QY 2192 ACAAGAGCGAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAAGG 2251
Db 2041 ACAGAGCGAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAAGG 2100
QY 2252 TGTAACCTGAGCTGGGTGCCCGCCACAAAGGGCATCGCGGGCAACGAGCAGATCGACAAGC 2311
Db 2101 TGTAACCTGAGCTGGGTGCCCGCCACAAAGGGCATCGCGGGCAACGAGCAGGTGGACAAGC 2160
QY 2312 TGGTAGCAGAGGAGATCGCGCAAGGTGCTGTTCTGAGAGGGCATCGATGGCGGCATCGTGA 2371
Db 2161 TGGTAGCAGCGCGGCATCGCGCAAGGTGCTGTTCTGAGAGGGCATCGATGGCGGCATCGTGA 2220
QY 2372 TCTACCACTACATGAGCAGCAGCTGAGTGGGAGCGGGCGGGCCCTAGGATCGATTAAAGC 2431
Db 2221 TCTACCACTACATGAGCAGCAGCTGAGTGGGAGCGGGCGGGCCCTAGGATCGATTAAAGC 2280
QY 2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457
Db 2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306

```

```

RESULT 3
US-09-475-515-84
; Sequence 84, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 84
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.Rtopt(+)
US-09-475-515-84

```

```

Query Match 81.3%; Score 1998.4; DB 4; Length 2312;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

```

```

QY 170 GCGCAAGGAGGCGCCACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCG 229
Db 1 GCGCGCGGAGAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60
QY 230 AGGACTTGGCTTCCCGCAAGGCGCGAGTTCCCGAGCGAGAGAACCGCGCA 289
Db 61 AGGACTTGGCTTCCCGCAAGGCGCGAGTTCCCGAGCGAGAGAACCGCGCA 120
QY 290 ACAGCCCCACAGCGCGAGCTGCAGTGGCGGGG-----ACAAACCCCGCAGCGAGG 343
Db 121 ACAGCCCCACCGCGCGAGCTGCAGTGGCGGGGCGGAGAACAAACAGCTTCAGCGAGG 180
QY 344 CCGCGCGAGCGCGCGAGGCGCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGC 397
Db 181 CCGCGCGCGAGCGCGCGAGGCGCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGC 240
QY 398 GCGCCCTTGGTGAAGCATCAAGGTGGCGCGCGAGATCAAGGAGGCGCTGTGGACACCGCG 457
Db 241 GCGCCCTTGGTGAAGCATCAAGGTGGCGCGCGAGCTCAAGGAGGCGCTGTGGACACCGCG 300
QY 458 CCGAGCACACCGTGTGAGAGAGATGAGCTGCGCGCGCAAGTGGAGGCCCAAGATGATCG 517
Db 301 CCGAGCACACCGTGTGAGAGAGATGAACTGCGCGCGCAAGTGGAGGCCCAAGATGATCG 360
QY 518 GCGGCATCGCGGGCTTCATCAAGGTGGCGCGAGTACGACAGATCCTGATCGAGATCTGCG 577
Db 361 GCGGCATCGCGGGCTTCATCAAGGTGGCGCGAGTACGACAGATCCTGATCGAGATCTGCG 420
QY 578 GCAAGAGGCCATCGGCACCGTGTGATCGGCGCCACCGCGCGCAACATCATCGCGCGCA 637
Db 421 GCAAGAGGCCATCGGCACCGTGTGATCGGCGCCACCGCGCGCAACATCATCGCGCGCA 480
QY 638 ACATGCTGACCGAGCTGGGCTGCACCTGAGACTTCCCATCAGCCCCCATCGAGACCGTGC 697
Db 481 ACCTGCTGACCGAGATCGGCTGCACCTGAACTTCCCATCAGCCCCCATCGAGACCGTGC 540
QY 698 CCGTGAAGCTGAAGCGCGCGCATGGAGCGCCCAAGGTGAAGCGTGGCGCCCTGACCGAGG 757
Db 541 CCGTGAAGCTGAAGCGCGGGATGGAGCGGCCCAAGGTGAAGCGTGGCGCCCTGACCGAGG 600
QY 758 AGAAGTCAAGGCGCTGACCGGCACTTCGCGAGGAGATGGAGAGGAGGCGCAAGATCAACA 817
Db 601 AGAAGTCAAGGCGCTGCTGGAGATCTGCACCGAGATGGAGAGGAGGCGCAAGATCAGCA 660
QY 818 AGATCGGCGCGAGAACCCCTACAACACCGCGCTGTTCCGCATCAAGAGAGGAGCAGCA 877
Db 661 AGATCGGCGCGAGAACCCCTACAACACCGCGCTGTTCCGCATCAAGAGAGGAGCAGCA 720
QY 878 CCAAGTGGCGCAAGCTGTGGACTTCCCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db 721 CCAAGTGGCGCAAGCTGTGGACTTCCCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780
QY 938 AGGTGAGCTGGGCATCCCGACCGCGCGCTGAAGAAGAGAGAGCGTGAACGCTGC 997
Db 781 AGGTGAGCTGGGCATCCCGACCGCGCGCTGAAGAAGAGAGAGCGTGAACGCTGC 840
QY 998 TGGAGCTGGGCGACGCTTACTTACGCTGCGCTGAGCGAGACTTCCGCAAGTACACCG 1057
Db 841 TGGAGCTGGGCGACGCTTACTTACGCTGCGCTGAGCGAGACTTCCGCAAGTACACCG 900
QY 1058 CCTTCACCATCCCGAGCATCAACAACGAGACCGCGGATCCGCTACAGTACAACAGCTGC 1117
Db 901 CCTTCACCATCCCGAGCATCAACAACGAGACCGCGGATCCGCTACAGTACAACAGCTGC 960
QY 1118 TGCCCCAGGCTGGAAGGCGCGCCAGCATCTTCAGAGCAGCATGACCAAGATCTGCG 1177
Db 961 TGCCCCAGGCTGGAAGGCGCGCCAGCATCTTCAGAGCAGCATGACCAAGATCTGCG 1020
QY 1178 AGCCCTTCCGCGCGCGCAACCGCGAGATCGTGAATCTACCA-----GGCCCCCTGTACG 1231
Db 1021 AGCCCTTCCGCGCGCGCAACCGCGAGATCGTGAATCTACCA-----GGCCCCCTGTACG 1080

```

Qy 1232 TGGGAGCGAGCTGTGAGATCGGCGAGCAACCGGCGCAAGATCGAGAGCTGCGCAAGCAC 1291  
Db 1081 TGGGAGCGAGCTGTGAGATCGGCGAGCAACCGGCGCAAGATCGAGAGCTGCGCGAGCAC 1140  
Qy 1292 TGCTGCGCTGGGCTTACACACCCCGGCAAGAGCACCAGAGAGCCCGCTTCTGTC 1351  
Db 1141 TGCTGCGCTGGGCTTACACACCCCGGCAAGAGCACCAGAGAGCCCGCTTCTGTC 1200  
Qy 1352 CCAT-----CGAGTGCACCCCGGCAAGAGTGGACCGTGCAGCCCATCGAGTGGCCGAGA 1405  
Db 1201 GGATGGGCTACGAGCTGCACCCCGGCAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGA 1260  
Qy 1406 AGGAGAGCTGACCGTGAACGACATCCAGAGCTGGTGGGCAAGTGAATGGGCGAGCC 1465  
Db 1261 AGGAGAGCTGACCGTGAACGACATCCAGAGCTGGTGGGCAAGTGAATGGGCGAGCC 1320  
Qy 1466 AGATCTACCCCGGATCAAGGTGGCCAGCTGTCAGAGTCTGCGGGCGCCAGAGCC 1525  
Db 1321 AGATCTACCCCGGATCAAGGTGAAGCTGTGCAAGCTGTGCGGGCACCAGAGCC 1380  
Qy 1526 TGACCGACATCGTGCCTGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCGAGA 1585  
Db 1381 TGACCGAGGTGATCCCTGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCGAGA 1440  
Qy 1586 TCCTGCGGAGCCGTGACCGGCGTGTACTAGACCCGAGCAAGACTGGTGGCCGAGA 1645  
Db 1441 TCCTGAAGGAGCCGTGACCGGCTGTACTAGACCCGAGCAAGACTGGTGGCCGAGA 1500  
Qy 1646 TCCAGAGAGCGGCCAGACGAGTGGACCTACGAGATCTACGAGAGCCCTTCAAGAAC 1705  
Db 1501 TCCAGAGAGCGGCCAGGCGGAGTGGACCTACGAGATCTACGAGAGCCCTTCAAGAAC 1560  
Qy 1706 TGAAGACCGGCAAGTACGCAAGATGCGACCGCCGACCAACGACGAGTGAAGCAGTGA 1765  
Db 1561 TGAAGACCGGCAAGTACGCGCGATGCGGCGGCCACCAACGAGCTGAAGCAGTGA 1620  
Qy 1766 CCGAGGCGGTGACAGAGATGCGCATGGAGAGCATGTGATCTGGGGCAAGACCCCAAGT 1825  
Db 1621 CCGAGGCGGTGACAGAGTGAACCGAGAGCATGTGATCTGGGGCAAGATCCCAAGT 1680  
Qy 1826 TCCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGAACCGACTACTGGGAGGCCA 1885  
Db 1681 TCAAGCTGCCCATCCAGAGGAGACCTGGGAGGCTGGTGAATGAGTACTGGGAGGCCA 1740  
Qy 1886 CCTGATCTCCGAGTGGGAGTTCGTGAACACCCCGCTGCTGTAAGCTGTGGTACAGC 1945  
Db 1741 CCTGATCTCCGAGTGGGAGTTCGTGAACACCCCGCTGCTGTAAGCTGTGGTACAGC 1800  
Qy 1946 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTTCTACGTGGAACGCGCGCCCAACCGCG 2005  
Db 1801 TGGAGAGGAGCCCATCGTGGGCGCGAGACCTTTCTACGTGGAACGCGCGCCCAACCGCG 1860  
Qy 2006 AGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGCGGAGAGATCGTGAGCC 2065  
Db 1861 AGACCAAGCTGGGCAAGCGCGCTACGTGACCGACCGGGCGCGGAGAGTGGTGAGCA 1920  
Qy 2066 TGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGGACA 2125  
Db 1921 TCGCCGACACCAACCAACAGAGACCGAGCTGCAGGCGCATCCACTGGCCCTGCAGGACA 1980  
Qy 2126 GCGCAGCGAGGTGAACAATCGTGAACGACAGCAGTACGCGCTGGGCGATCATCCAGGCCC 2185  
Db 1981 GCGCCTGGAGGTGAACAATCGTGAACGACAGCAGTACGCGCTGGGCGATCATCCAGGCCC 2040  
Qy 2186 AGCCCGACAGAGCGAGCGAGCTGGTGAACCAAGATCATGAGCAGCTGATCAAGAGG 2245  
Db 2041 AGCCCGACAGAGCGAGCGAGCTGGTGAACCAAGATCATGAGCAGCTGATCAAGAGG 2100  
Qy 2246 AGAAGGTGATCTGAGCTGGTGGCCCGCCAGAGGCGATCGGGCGGCAAGCGAGATCG 2305  
Db 2101 AGAAGGTGATCTGAGCTGGTGGCCCGCCAGAGGCGATCGGGCGGCAAGCGAGATGG 2160  
Qy 2306 ACAAGCTGGTGAAGAGGCGATCCGCAAGGTGCTGTTCTTGGACGGCATCGATGGCGGCA 2365

Db 2161 ACAAGCTGGTGAAGCGCGGATCCGCAAGGTGCTGTTCTGAAACGGCATCGATGGCGCA 2220  
Qy 2366 TCGTGATCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 2425  
Db 2221 TCGTGATCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 2280  
Qy 2426 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

## RESULT 4

US-09-475-515-6  
; Sequence 6, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475.515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 4319  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-gag-polymerase  
US-09-475-515-6

Query Match 78.0%; Score 1915.8; DB 4; Length 4319;  
Best Local Similarity 89.4%; Pred. No. 0;  
Matches 2140; Conservative 0; Mismatches 227; Indels 28; Gaps 6;

Qy 14 TGGCGAGGCGCATGAGCCAGGCGCACAGC---GCCAACATCTGATGAGCGCAGCAACT 70  
Db 1100 TGGCGAGGCGCATGAGCCAGGCGCACAGC---GCCAACATCTGATGAGCGCAGCAACT 1159  
Qy 71 TCAAGGCGCCCAAGCGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGCGCCACATCGGCC 130  
Db 1160 TCGCAACAGCGGAGAACCGTCAAGTGTCTCAACTCGGCAAGGAGGCGCCACATCGGCC 1219  
Qy 131 GCAATCGCCGCGCCCGCCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCCACCA 190  
Db 1220 GGAATCGCCGCGCCCGCCGCAAGAGGCTGCTGCGCTGCGGCGCGGAGGACACCAA 1279  
Qy 191 TGAAGGACTGACGAGCGCGCAGGCGCAACTTTCTTCGCGAGGAGCTGCGCTTCCCGCAGG 250  
Db 1280 TGAAGATTCCTGACTGAGAGACAGGCTAA-TTTTATTAGGGAAGATCTGCGCTTCTTACAAG 1338  
Qy 251 GCAAGGCGCGGAGTTCCTCCAGCGAGCAGACCGCGCAACAGCGCCACAGCGCGGAGC 310  
Db 1339 GGAAGGCGCAGGGAATTTTCTTTCAGAGCAGACCGAGCGCAACAGCGCCACAGAGAGAGC 1398  
Qy 311 TGCAGGTGCGCGCGG-----ACAAACCCCGCAGCGAGGCGCGCGCGAGCGCGCAGGCGCA 364  
Db 1399 TGCAGGTGCGCGCGGAGGAGAAACAACTCTCTCAGAGCAGGAGCGCGATAGACAGGAA 1458  
Qy 365 -----CCCTGAACTTCCCGCAGATCACTCTGTGGCAGCGCCCTCGTGAGCATCAAGG 418  
Db 1459 CTGTATCTCTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACTAGTAAGGA 1518

QY 419 TGGCGGCGCAGATCAAGGAGGCGCTGCTGGACACCGGCGCCGACGACACCGTGTGGAGG 478  
 Db 1519 TCGGCGGCGCAGCTCAAGGAGGCGCTGCTCGACACCGGCGCCGACGACACCGTGTGGAGG 1578  
 QY 479 AGATGAGCCTGCGCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 538  
 Db 1579 AGATGAACCTGCGCGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 1638  
 QY 539 AGGTGGCCAGTACGACCGAGATCTGTATCGAGATCTGGCGGCAAGAGGCGCATCGGACCG 598  
 Db 1639 AGGTGGCGAGTACGACCGAGATCTGTATCGAGATCTGGCGGCAAGAGGCGCATCGGACCG 1698  
 QY 599 TGCTGATCGGCGGCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 658  
 Db 1699 TGCTGATCGGCGGCGGCGGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA 1758  
 QY 659 GCACCTGAACTTCCGCGCATCGAGCCCGCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCA 718  
 Db 1759 GCACCTGAACTTCCGCGCATCGAGCCCGCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCA 1818  
 QY 719 TGGACGCGCCCAAGTGAAGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 778  
 Db 1819 TGGACGCGCCCAAGTGAAGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1878  
 QY 779 CCATCTGCGAGGAGATGGAGAGGAGGCGGCAAGATCAACCAAGATCGGCGGCGGCGGCGGCGG 838  
 Db 1879 AGATCTGCAACCGAGATGGAGAGGAGGCGGCAAGATCAGCAAGATCGGCGGCGGCGGCGGCGG 1938  
 QY 839 ACAACACCCCGGCTGTCGCGCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 898  
 Db 1939 ACAACACCCCGGCTGTCGCGCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 958  
 QY 899 ACTTCGCGAGCTGAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 958  
 Db 1999 ACTTCGCGAGCTGAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2058  
 QY 959 ACCCGCGCGGCTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1018  
 Db 2059 ACCCGCGCGGCTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2118  
 QY 1019 TCAGGCTGCGGCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1078  
 Db 2119 TCAGGCTGCGGCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2178  
 QY 1079 ACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1138  
 Db 2179 ACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2238  
 QY 1139 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1198  
 Db 2239 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2298  
 QY 1199 CCGAGATCGTGTATCAACA-----GGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1252  
 Db 2299 CCGAGATCGTGTATCAACA-----GGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2358  
 QY 1253 GCGAGAG 1312  
 Db 2359 GCGAGAG 2418  
 QY 1313 CCGCGGAG 1366  
 Db 2419 CCGCGGAG 2478  
 QY 1367 CCGAGAG 1426  
 Db 2479 CCGAGAG 2538  
 QY 1427 ACATCAGAG 1486  
 Db 2539 ACATCAGAG 2598

QY 1487 TGGCGGAGCTGTGAAGCTGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1546  
 Db 2599 TGAAGCAGCTGTGCAAGCTGTGCGCGGCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2658  
 QY 1547 CCGAGAGGCGGAGCTGTGAGCTGTGCGGAGAACCGGAGATCTTGTGCGGAGGCGGCGGCGGCGG 1606  
 Db 2659 CCGAGAGGCGGAGCTGTGAGCTGTGCGGAGAACCGGAGATCTTGTGCGGAGGCGGCGGCGGCGG 2718  
 QY 1607 GCGTGTACTAGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1666  
 Db 2719 AGGTGTACTAGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2778  
 QY 1667 AGTGAGCTTACAGATCTACAGGAGGCGCTTCAAGAACCTGAAGACCGGCGGCGGCGGCGGCGG 1726  
 Db 2779 AGTGAGCTTACAGATCTACAGGAGGCGCTTCAAGAACCTGAAGACCGGCGGCGGCGGCGGCGG 2838  
 QY 1727 AGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1786  
 Db 2839 GCATGCG 2898  
 QY 1787 CCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1846  
 Db 2899 GCACCGAG 2958  
 QY 1847 AGACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1906  
 Db 2959 AGACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3018  
 QY 1907 TCGTGAACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1966  
 Db 3019 TCGTGAACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3078  
 QY 1967 GCGCGGAG 2026  
 Db 3079 GCGCGGAG 3138  
 QY 2027 GCTAGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2086  
 Db 3139 GCTAGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3198  
 QY 2087 AGACCGAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2146  
 Db 3199 AGACCGAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3258  
 QY 2147 TGAACCGAG 2206  
 Db 3259 TGAACCGAG 3318  
 QY 2207 AGCTGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2266  
 Db 3319 AGCTGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3378  
 QY 2267 TGCGCGG 2326  
 Db 3379 TGCGCGG 3438  
 QY 2327 TCCGAGAG 2381  
 Db 3439 TCCGAGAG 3493

RESULT 5

US-09-475-515-81  
 ; Sequence 81, Application US/09475515A  
 ; Patent No. 6602705  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BARNETT, Susan  
 ; APPLICANT: ZUR MEDEDE, Jan  
 ; APPLICANT: SRIVASTAVA, Indresh  
 ; APPLICANT: LIAN, Ying  
 ; APPLICANT: HARTOG, Karlin  
 ; APPLICANT: LIU, Hong  
 ; APPLICANT: GREER, Catherine

; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 81  
; LENGTH: 2299  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(+) proinact.RTopt.YMMW  
US-09-475-515-81

Query Match 76.2%; Score 1872.8; DB 4; Length 2299;  
Best Local Similarity 89.3%; Pred. No. 2.2e-310;  
Matches 2055; Conservative 0; Mismatches 232; Indels 13; Gaps 3;  
Qy 170 GCGCAAGGAGGCGCCACAGATGAAGGACTGCACCGAGGCGCCAACTTCTTCGCG 229  
Db 1 CGGCGCGGAGGACACCAATGAAGATTGCACTGAGAGAGCGCTAATTTTT-AGGG 59  
Qy 230 AGGACCTTGGCTTCCCCAGGCGCAAGGCCGCGAGTTCCCCAGCGAGCAGAAACCGCGCA 289  
Db 60 AAGATCTGGCTTCTTACAAGGGAAGGCCAGGGAATTTTCTTACAGAGCAGAGGCA 119  
Qy 290 ACAGCCCCACAGCGCGAGCTGCAAGTGCAGTGCAGGCGGGG-----ACAAACCCCGAGCGAGG 343  
Db 120 ACAGCCCCACAGAGAGAGCTTCAAGTTCAGGTTTGGGAGGAGAAACAACTCCCTCTCAGAG 179  
Qy 344 CCGCGCGGAGCGCCAGGCA-----CCCTGAACTTCCCCAGATCACCTGTGGGAGC 397  
Db 180 CAGAGCGGATAGACAGGAATCTGATCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239  
Qy 398 GCCCCTCGTGAGCATCAAGTGGCGGCGCCAGATCAAGGAGGCCCTGTGACACCGCG 457  
Db 240 GACCCCTCGTCAATAAGATCGGGGGCACTCAGGAAGCGTGTCTCATACAGGAG 299  
Qy 458 CCGAGCACCGTGTGAGAGATGAGCTTGCCTGGCAAGTGAAGCCCAAGATGATCG 517  
Db 300 CAGATGATACAGTATTAGAAAGATGAATTTGCCAGGAAATGGAACCAAAATGATAG 359  
Qy 518 GCGCATCGCGGCTTCAATCAAGTGCAGGAGTGCAGTACGACCATCTGATCAGATCTGCG 577  
Db 360 GGGGATCGGGGCTTCAATCAAGTGAAGGAGTACGACCATACCTGTAGAAATCTGTG 419  
Qy 578 GCAAGAGGCGCATCGGCACCGTGTGATCGGCGCCACCCCGTGAAATCATCATCGCGCA 637  
Db 420 GACATAAGCTATAGTACAGTATTAGTAGACCTACACCTGTCAATATTTGAAGAA 479  
Qy 638 ACATGCTGACCCAGTGGGCTGCAACCTTGAATCTTCCCATCAGCCCATCGAGACCGTGC 697  
Db 480 ATCTGTTGACCCAGATCGGCTGCACTTGAATCTTCCCATCAGCCCATATTGAGAGCGTGC 539  
Qy 698 CCGTGAAGCTCAAGCCCGCATGAGCGCCCAAGTGAAGCTGAGCGCCCTGACCGAGG 757  
Db 540 CCGTGAAGTGAAGCCGGGATGAGCGGCCCAAGGTCAAGCAATGCGCAATTGACCGAGG 599  
Qy 758 AGAAGATCAAGGCTTGAACCGCCATCTGAGGAGATGAGGAAGGAGGCGCAAGATCACCA 817  
Db 600 AGAAGATCAAGGCTTGTGAGATCTGCAACCGAGATGAGGAAGGCGCAAGATCAGCA 659  
Qy 818 AGATCGGCGCGAGAAACCCCTTAAACACCCCGTGTGCGCATCAAGAAGAGACAGCA 877  
Db 660 AGATCGGCGCGAGAAACCCCTTAAACACCCCGTGTGCGCATCAAGAAGAGACAGCA 719  
Qy 878 CCAAGTGGCGAGCTGTGAGTCTTCCGAGCTGAAACAGCGCACCCAGACTTCTGGG 937  
Db 720 CCAAGTGGCGAGCTGTGAGTCTTCCGAGCTGAAACAGCGCACCCAGACTTCTGGG 779

Qy 938 AGGTGCAGCTGGGCATCCCCACCCCGCGGCTGAAGAAGAAGAGCGTGAACGTGC 997  
Db 780 AGGTGCAGCTGGGCATCCCCACCCCGCGGCTGAAGAAGAAGAGCGTGAACGTGC 839  
Qy 998 TGGAGTGGGCGAGCGCTTACTTACAGTGGCCCTGGAGAGAGATTCGCAAGTACACG 1057  
Db 840 TGGAGTGGGCGAGCGCTTACTTACAGTGGCCCTGGAGAGAGATTCGCAAGTACACG 899  
Qy 1058 CTTTACCATCCAGCATCAACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117  
Db 900 CTTTACCATCCAGCATCAACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 959  
Qy 1118 TGCCCCAGGCTGGAAGGCGAGCGCCAGCATCTTTCAGAGCAGCATGACCAAGATCTCTGG 1177  
Db 960 TGCCCCAGGCTGGAAGGCGAGCGCCAGCATCTTTCAGAGCAGCATGACCAAGATCTCTGG 1019  
Qy 1178 AGCCCTTCCGCGCGCGCAACCCCGAGATCGTGATCTTACAGAGCGCCCTTACTAGTGGCA 1237  
Db 1020 AGCCCTTCCGCAAGCAGAAACCCCGCATCTGATCTTACAGAGCGCCCTTACTAGTGGCA 1079  
Qy 1238 GCGACCTGGAGATCGGCGAGCAGCGCGCAAGATCGAGAGCTGCGCAAGCAGCATCTGTCG 1297  
Db 1080 GCGACCTGGAGATCGGCGAGCAGCGCGCAAGATCGAGAGCTGCGCAAGCAGCATCTGTCG 1139  
Qy 1298 GCTGGGCTTCCACCAACCCCGAGAGAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAG 1357  
Db 1140 GCTGGGCTTCCACCAACCCCGAGAGAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAG 1199  
Qy 1358 AGCTGCACCCCGAGCAAGTGGAGCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGAGAG 1417  
Db 1200 AGCTGCACCCCGAGCAAGTGGAGCGTGCAGCGCATCTGCTGCGCGAGAGAGAGAGAG 1259  
Qy 1418 CCGTGAAGCAGCATCCAGAGCTGTGGCAAGCTGAACTGGGCGAGCAGAGATCTACCCCG 1477  
Db 1260 CCGTGAAGCAGCATCCAGAGCTGTGGCAAGCTGAACTGGGCGAGCAGAGATCTACCCCG 1319  
Qy 1478 GCATCAAGGTGGCGAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCCCTGACCGAGATCG 1537  
Db 1320 GCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCACCAAGGCCCTGACCGAGAG 1379  
Qy 1538 TGCCCTTGAACGAGAGAGCGAGCTGAGAGTGGCGCGAGAAACCGGAGAGATCTTGGCGAGC 1597  
Db 1380 TCCCTTGAACGAGAGAGCGAGCTGAGAGTGGCGCGAGAAACCGGAGAGATCTTGAAGAGC 1439  
Qy 1598 CCGTGCAGCGCTGTACTAGACCCAGAGAGAGCTGTGCGCGAGAGATCCAGAGAGAG 1657  
Db 1440 CCGTGCAGCGCTGTACTAGACCCAGAGAGAGCTGTGCGCGAGAGATCCAGAGAGAG 1499  
Qy 1658 GCCACGACAGTGGACCTTACAGATCTACAGAGAGCGCTTCAAGAACTGAGAGAGCGGCA 1717  
Db 1500 GCCAGGCGCAGTGGACCTTACAGATCTACAGAGAGCGCTTCAAGAACTGAGAGAGCGGCA 1559  
Qy 1718 AGTACGCGCAGATGCGCACCGCCCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1777  
Db 1560 AGTACGCGCAGATGCGCGCGCGCCCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1619  
Qy 1778 AGAAGATGCGCATCGAGAGAGATCGTGATCTGGGCGAGAGAGAGAGAGAGAGAGAGAG 1837  
Db 1620 AGAAGATGAGCACCGAGAGAGATCGTGATCTGGGCGAGAGATCCCCAAGTTCAGCTGCGCA 1679  
Qy 1838 TCCAG 1897  
Db 1680 TCCAG 1739  
Qy 1898 AGTGGAGAGTGGTGAACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1957  
Db 1740 AGTGGAGAGTGGTGAACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1799  
Qy 1958 CCATCATCGGCGCGAG 2017  
Db 1800 CCATCATCGGCGCGAG 1859





```
Qy 1356 ----CGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCCGAGAAGGAGA 1411
Db 1200 GCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACA 1259
Qy 1412 GCTGGACCGGTGAACGACATCCAGAGCTGTTGGGCAAGCTGTAACCTGGGCCAGCCAGATCT 1471
Db 1260 GCTGGACCGGTGAACGACATCCAGAGCTGTTGGGCAAGCTGTAACCTGGGCCAGCCAGATCT 1319
Qy 1472 ACCCGGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGGGCGCCCAAGGCCCTGACCG 1531
Db 1320 AGCGCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGGGCGCCCAAGGCCCTGACCG 1379
Qy 1532 ACATGTGTCCTTACCCGAGGAGCCGAGCTGAGCTGTGCGGAGAACCCGCGAGATCTCTGC 1591
Db 1380 AGGTGATCCCTTACCCGAGGAGCCGAGCTGTGAGCTGTGCGGAGAACCCGCGAGATCTCTGA 1439
Qy 1592 GCGAGCCGTGTACCGGGTGTACTACGACCCACAGAGACCTGTGTGCGCGAGATCCAGA 1651
Db 1440 AGGAGCCGTGTACCGGGTGTACTACGACCCACAGAGACCTGTGTGCGCGAGATCCAGA 1499
Qy 1652 AGCAGGCCACGACGAGTGAACCTTACGAGATCTACAGGAGCCCTTCAAGAACCTGAAGA 1711
Db 1500 AGCAGGCCACGAGCTGACCTTACGAGATCTACAGGAGCCCTTCAAGAACCTGAAGA 1559
Qy 1712 CCGCAAGTACGCCAAGATCGCACCGCCACACCAACGAGCTGAAGCAGCTGACCCGAGG 1771
Db 1560 CCGCAAGTACGCCAAGATCGCGCGCGCCACACCAACGAGCTGAAGCAGCTGACCCGAGG 1619
Qy 1772 CCGTGCAGAGATCGCCATGAGAGAGATCGTGATCTGTGGGGCAAGACCCCAAGTTCGCGC 1831
Db 1620 CCGTGCAGAGAGTGAACACCGAGAGCATCGTGATCTGTGGGGCAAGATCCCAAGTTCGAAGC 1679
Qy 1832 TGCCCATCCAGAGGAGACCTTGGGAGACCTGTGTGACCGGACTACTGTGCAGGCCACCTGGA 1891
Db 1680 TGCCCATCCAGAGGAGACCTTGGGAGCCCTGTGTGATGGAGTACTTGGGAGGCCACCTGGA 1739
Qy 1892 TCCCAGAGTGGGAGTTCGTGAACACCCGCCCTTGTGTGAAGCTGTGTGTACAGCTGGAGA 1951
Db 1740 TCCCAGAGTGGGAGTTCGTGAACACCCGCCCTTGTGTGAAGCTGTGTGTACAGCTGGAGA 1799
Qy 1952 AGAGGCCATCATCGGCGCCGAGACCTTCTACGTGGAAGCTGTGTGTACAGCTGGAGACCA 2011
Db 1800 AGAGGCCATCATCGGCGCCGAGACCTTCTACGTGGAAGCTGTGTGTACAGCTGGAGACCA 1859
Qy 2012 AGATCGGCAAGGCCGCTACGTGACCGACCGGGGCCGCGAGAGATCGTGAGCTGACCG 2071
Db 1860 AGTGGCAAGGCCGCTACGTGACCGACCGGGGCCGCGAGAGAGGTGTGAGCATCGCCG 1919
Qy 2072 AGACCAACCAAGAGACCGAGCTGCAGGCCATCTCAGCTGGCCCTGCAGGACAGCGGCA 2131
Db 1920 ACACCAACCAAGAGACCGAGCTGCAGGCCATCTCAGCTGGCCCTGCAGGACAGCGGCC 1979
Qy 2132 GCGAGGTGAACATCGTGACCGACAGCAGTACGCGCTTGGGATCATCAGGCCCGAGCCCG 2191
Db 1980 TGGAGGTGAACATCGTGACCGACAGCAGTACGCGCTTGGGATCATCAGGCCCGAGCCCG 2039
Qy 2192 ACAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2251
Db 2040 ACAAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2099
Qy 2252 TGTACCTGTAGCTGGGTGCCCGCCACCAAGGCAATCGCGGCAACGAGCAGATCGCAAGC 2311
Db 2100 TGTACCTGTAGCTGGGTGCCCGCCACCAAGGCAATCGCGGCAACGAGCAGATCGCAAGC 2159
Qy 2312 TGTGAGCAAGGGCATCCGCAAGTGTGTGTCTTGGACGGCATCGATGGCGGCATCGTGA 2371
Db 2160 TGTGAGCGCGGCATCCGCAAGTGTGTGTCTTGAACGGCATCGATGGCGGCATCGTGA 2219
Qy 2372 TCTACCAAGTACATGGACGACCTGTACGTGGGACGCGGCCCTTAGATCGATTAAAGC 2431
Db 2220 TCTACCAAGTACATGGACGACCTGTACGTGGGACGCGGCCCTTAGATCGATTAAAGC 2279
```

```
Qy 2432 TTCCCGGGCTAGCACCGGTGAATTC 2457
Db 2280 TTCCCGGGCTAGCACCGGTGAATTC 2305

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552.950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-synnp - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

Query Match 66.2%; Score 1626.6; DB 4; Length 4307;
Best Local Similarity 82.2%; Pred. No. 2.le-268;
Matches 1951; Conservative 0; Mismatches 394; Indels 28; Gaps 6;

Qy 12 CATGGCGGAGCCATGAGCCAGG---CCACGAGCGCCAAATCTCTGATGACGCGGAGCAA 68
Db 1086 CTTGGCTGAGGCCATGAGCCAGGTCGACCACTCCGCTACCATCATGATGACGCGGCGCAA 1145
Qy 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGCCACATCGC 128
Db 1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTCGCAAGAGAGGCCACACAGC 1205
Qy 129 CCGCAACTGCGGCGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAGGCCACCA 188
Db 1206 CCGCAACTGAGGGCCCTTAGGAAAGGGCTGTGGAATTCGCGCAAGGAGGCCACCA 1265
Qy 189 GATGAAGGACTGACCAAGCGCCAGGCAACTTCTTCCGCGAGGACCTTGGCTTCCCGCA 248
Db 1266 GATGAAGGACTGTACGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCTTCTTACA 1324
Qy 249 GGGCAAGGCCCGGAGTTCCTCCAGCGAGCAAGACCGCGCCCAACAGCCCGCCAGCGGGA 308
Db 1325 AGGGAAGGCCAGGGAATTTTCTTCAAGAGCAGAGACCGCGCCCAACAGCCCGCGCA 1384
Qy 309 GCTGCAAGT-----GCGCGCGACACACCCCGCAGCGAGGCGCGCGCGAGCCAGGG 362
Db 1385 GCTTCAAGTCTGGGGTGCAGCAACAACTCCCCCTCCGAAGAGGAGCGCGAGCCAGGG 1444
Qy 363 CA-----CCCTGAACATTTCCCGAGATCACCTGTGCGAGCGCCCTTGTGTGAGCATCAA 416
Db 1445 CACGCTGTCTTCAACTTCTTCCCTCAGGTCAAGCTTGGCAGGACCCCTCGTACCATCAA 1504
Qy 417 GGTGGGCGGCAGATCAAGAGGCGCTGTGTGACACCGCGCGCGCAACAGACCGTGTGTGA 476
Db 1505 GATCGGGGGCAGCTCAAGGAGGCTCTCTCTGGAACACCGGAGCAGACGACCGTGTGTGA 1564
Qy 477 GGAATGAGGCTCCCGCGCAAGTGGAGCCCAAGATGATCGGGGCAATCGCGGCTTCAT 536
Db 1565 GGAATGATGCTGTCAGAGCCGCTGGAAGCCGAAGATGATCGGGGGAATCGCGGCTTCAT 1624
Qy 537 CAAGGTGCGCCAGTACAGCCAGATCTCTGATCGAGATCTGGCGCAAGAGCCATCGGCGAC 596
Db 1625 CAAGGTGCGCCAGTATGACCAAGATCTCTCATCGAAATCTGGCGCAACAGGCTATCGGTAC 1684
Qy 597 CGTGTGATCGGCGCCACACCCCGTGAAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGG 656
Db 1685 CGTGTGATCGGCGCCACACACCCCGTGAAACATCATCGGAGCGCAACCTGTGTGACGAGATCGG 1744
```

QY 657 CTGCACTTCCCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCCCGG 716  
DB 1745 TTGCACGCTGAATCTTCCCATTAGCCCTATCAGACCGGTACCGGTGAAGCCCGG 1804  
QY 717 CATGACCGCCCAAGAGTGAACAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 776  
DB 1805 GATGACCGCCCAAGAGTGAACAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGAC 1864  
QY 777 CGCCATCTGCGAGGAGATGGAGAGGAGGCAAGATCAACAGATCGGCCCGCGAAGCC 836  
DB 1865 GGAGATTTGCA CAGAGATGGAAAAGGAGGAAAATCTCCAAGATTGGCGCTGAGAACCC 1924  
QY 837 CTACACACCCCGCTTCCCATCAAGAGAGGAGCAGCAGCAGTGGCGCAAGCTGGT 896  
DB 1925 GTACACACCGCGGTGTTCCCATCAAGAGAGGAGCAGCAGTGGCGCAAGCTGGT 1984  
QY 897 GGACTTCCCGGAGCTGAACAAGCAGCAGCAGCAGTCTTGGAGGTGCAAGTGGGATGCC 956  
DB 1985 GGACTTCCCGGAGCTGAACAAGCAGCAGCAGCAGTCTTGGAGGTGCAAGTGGGATGCC 2044  
QY 957 CCACCCCGCGCTTGAAGAGAGAGAGCGTGAACCGTGTGGAAGTGGGCGAGCCTTA 1016  
DB 2045 GCACCCCGCAGGCTTGAAGAGAGAGAGTCCGTGACCGCTACTGGATGTGGGTGATGCTTA 2104  
QY 1017 CTTAGCGTCCCTGGAGGAGCTTCCGCAAGTACACCGCTTCAACATCCCTCAGCAT 1076  
DB 2105 CTTCTCCGTTCCCTGGAGGAGCTTCCGCAAGTACACCGCTTCAACATCCCTCAGCAT 2164  
QY 1077 CAACACAGAGACCCCGGATCCGTACAGTACCAAGTGTGCTGCGCCCGGCTGGAAGG 1136  
DB 2165 CAACACAGAGACCCCGGATTCGATATCAGTACAAAGTGTGCTGCGCCCGGCTGGAAGG 2224  
QY 1137 CAGCCCGAGCTTCCAGAGAGAGATGACCAAGATCTCTGAGAGCCTTCCCGCCCGCAA 1196  
DB 2225 CTTCTCCGCAATCTTCCAGAGTGAATGACCAAAATCTCTGAGCCTTCCCGCAACAGAA 2284  
QY 1197 CCGCAGATCGTATCTACCA-----GGCCCCCTGTACGTGGCAGGACCTGGAGAT 1250  
DB 2285 CCGCAGATCGTATCTACCA-----GGCCCCCTGTACGTGGCAGGACCTGGAGAT 2344  
QY 1251 CCGCAGACCCCGCAGAGCTGAGGAGCTGCGCAAGCAGCAGTGTGCGTGGGGCTTCAAC 1310  
DB 2345 AGGGCAGACCCGACCAAGATCGAGGAGCTGCGCAGCAGCAGTGTGAGTGGGAGTGAAC 2404  
QY 1311 CACCCCGACAGAGAGCAGAGAGAGCCCGCTTCTGCCCCAT-----CGAGCTGCA 1364  
DB 2405 CACACCCGACAGAGAGCAGAGAGAGCCCGCTTCTGCCCCAT-----CGAGCTGCA 2464  
QY 1365 CCGCAGAGTGGACCGTGCAGCCCATCGAGCTGCGCCGAGAGGAGCTGGACCGTGA 1424  
DB 2465 CCGTCAAAATGGACCGTGCAGCCCTATCGTGTGCGCAGAAAGAGAGCTGGAGCTGCA 2524  
QY 1425 CGACATCCAGAGCTGGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTACCCCGCATCAA 1484  
DB 2525 CGACATACAGAGCTGGTGGGCAAGTGAATCTGGGCGCAGTCAATTTACCCAGGATTA 2584  
QY 1485 GGTGCGCAGCTGTGAAGCTGCTGCGCGCGCCAGGCGCTTGAACGAGATCGTGGCCCT 1544  
DB 2585 GGTGAGGAGCTGTGCAAACTCTCCGCGGAACCAAGGCACTCAGAGAGTGTATCCCT 2644  
QY 1545 GACCGAGGAGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTTCCGCGAGCCGTGCA 1604  
DB 2645 AACCGAGGAGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTTAAAGGAGCCGTGCA 2704  
QY 1605 CGCGCTGTACTACGACCCCGCAGCAGGACCTGTGTGCGCGAGATCCAGAGCAGGCGCA 1664  
DB 2705 CGCGCTGTACTATGACCCCTTCAAGGAGCTGTATCGCGAGATCCAGAGCAGGCGCA 2764  
QY 1665 CAGTGGAGCTTACAGAGCTTACAGAGCGCTTCAAGAACCTTGAAGAGCCGGAAGTACGC 1724  
DB 2765 CCAGTGGAGCTTACAGATTTTACAGAGCGCTTCAAGAACCTTGAAGAGCCGGAAGTACGC 2824

QY 1725 CAAGATCGCACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAAGAT 1784  
DB 2825 CCGATGAGGGTGGCCACACTTAACGACTCAGCAGCTGACCGAGGCGGTGCAAGAT 2884  
QY 1785 CGCCATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAA 1844  
DB 2885 CACCACCAAGAGCATCTGTGATCTGGGAAAGACTCCTAAGTTCAAGCTGCCCATCCAGAA 2944  
QY 1845 GGAGACCTGGAGACCTGTGTGACCGGACTACTGGCAGGCGCCTGGATCCCGAGTGGGA 1904  
DB 2945 GGAACCTGGGAAACCTGTGTGACAGAGTATTGGCAGGCGCCTGGATTTCTGATGGGA 3004  
QY 1905 GTTCTGTAAACCCCGCCCTGTGTGAAGCTGTGTGATACCGCTGGAGAGGAGCCCATCAT 1964  
DB 3005 GTTCTGTAAACCCCGCCCTGTGTGAAGCTGTGTGATACCGCTGGAGAGGAGCCCATCAT 3064  
QY 1965 CGCGCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2024  
DB 3065 GGGCGCCGAAACCTTCTACGTGGATGGGCGCGCTAAACAGGGAGACTAAGCTGGGCAAGC 3124  
QY 2025 CGCTTACGTGACCGACCGGGCGCGCAGAGATCGTGAAGCTGACCGGAGACCAACCAACA 2084  
DB 3125 CGGATACGTCTAAACCGGGCGCAGACAGAGGTTGTCAACCTCACTGACACCAACCAACA 3184  
QY 2085 GAAGACCGAGCTGCAAGGCGCATCCAGCTGGCCCTGCGAGCAGCGGCGAGCTGAACAT 2144  
DB 3185 GAAGACTGAGCTGCAAGGCGCATTTACCTGCTTTGCAAGACTCGGGCTGGAGTGAACAT 3244  
QY 2145 CGTGACCAAGCCAGTACGCTGGGCGATCATCCAGGCGCCAGCCGCAAGAGCGAGAG 2204  
DB 3245 CGTGACGACTCTCAGTATGCTGGCGCATCATTAAGCCAGCCAGCAGAGTGAAGTC 3304  
QY 2205 CGAGCTGTGAACAGGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACTGAGCTG 2264  
DB 3305 CGAGCTGTGTAATCAGATCATCGAGCAGCTGATCAAGAGGAGGAGTGTACTGAGCTG 3364  
QY 2265 GGTGCCCCCAGAGGCGCATCGGCGCAACGAGCAGATCGACAGCTGGTGAAGAGG 2324  
DB 3365 GGTACCCCGCCAGAGGCGCATTTGGCGCATGAGCGAGTGTCAACAGCTGTCTCGGCTG 3424  
QY 2325 CATCGCAAGTGTCTTCTTCTGAGCGGATCGA 2357  
DB 3425 CATCAGGAAGTGTCTTCTTCTGAGTGTGATCGA 3457

## RESULT 8

US-09-936-572-2  
; Sequence 2, Application US/09936572  
; Patent No. 6783981  
; GENERAL INFORMATION:  
; APPLICANT: UDEN, MARK  
; APPLICANT: MITROPHANOUS, KYRIACOS  
; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
; FILE REFERENCE: 078883/0137  
; CURRENT APPLICATION NUMBER: US/09/936,572  
; CURRENT FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: GB 9906177.2  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4307  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: gapcol-SINgp-codon optimised gapcol sequence

Query Match 65.4%; Score 1607.4; DB 4; Length 4307;  
Best Local Similarity 81.7%; Pred. No. 3.9e-265;

Matches 1939; Conservative 0; Mismatches 406; Indels 28; Gaps 6;

```
Qy 12 CATGCCAGGCCATGAGCCAGG---CCACAGCGCCAAACATCTGTATGACGCGCAGCAA 68
Db
Qy 1086 CTTGGCTGAGGCCATGAGCCAGGTGACCAATCCGCTACCATCATGATGACGCGCGCAA 1145
Db
Qy 69 CTTTCAGGGCCCCAAGCGCATCATCAAGTGTTTCAACTGTGGGCAAGAGGGCCATCGC 128
Db
Qy 1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGTTTCAACTGTGGCAAGAGGGCACACAGC 1205
Db
Qy 129 CCGCAACTGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGSCCAACCA 188
Db
Qy 1206 CCGCACTGCAAGGCCCTTAGGAAGAGGGCTGTGGAATGCGGCAAGAGGCCACCA 1265
Db
Qy 189 GATGAAGACTGCAACGAGCGCCAGGGCCAACTTCTCCGCGAGGACCTGGCCTTCCCCCA 248
Db
Qy 1266 GATGAAGACTGTACTGAGAGACAGGCTAA-TTTTATTAGGAAGATCTGGCCTTCTTACA 1324
Db
Qy 249 GGGCAAGCGCGGAGTTCCCGACGAGCAGAACCGCGGCCAAAGCCCGCACCGCCGGA 308
Db
Qy 1325 AGGGAAGGCCAGGGAAATTTCTTCAGAGCAGACACAGAGCCCAACAGCCCGCCAGAAGA 1384
Db
Qy 309 GCTGCAGGTGCGGG-----CGACAACCCCGCAGCGAGGCGCGCGCGAGGCCCAGGG 362
Db
Qy 1385 GCTTCAGGTCTGGGGTAGAGAACAACTCCCTCTCAGAAAGCAGAGCCGATAGCAAGG 1444
Db
Qy 363 CA-----CCCTGAATTTCCCCAGATCACCTGTGGCAGCGCCCTCTGTGTGAGCATCAA 416
Db
Qy 1445 AACTGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAAGACCCCTCGTCAATATA 1504
Db
Qy 417 GTTGGGCGCCGAGATCAAGAGAGGCCCTGTGTGACACCGCGGCCGACAGACACCGTGTGGA 476
Db
Qy 1505 GATAGGGGGGCGAGCTCAAGAGGGCTCTCTTGGAACACCGGAGCAGACACCGTGTGGA 1564
Db
Qy 477 GGAGATGAGCTCGCCGCGCAAGTGGAGCCCAAGATGATCGGCGGCATCGCGCGCTTCAT 536
Db
Qy 1565 GGAGATGCTGTGTCAGCGCGCTGGAAGCCGAAGATGATCGGGGGAAATCGCGCGTTTCAT 1624
Db
Qy 537 CAAGGTGCGGCAGTACGACACAGATCTCTGATCGAGATCTGGGGCAAGAGGCCATCGGCAC 596
Db
Qy 1625 CAAGGTGCGGCAGTATGACGAGTCTCTATCGAATCTGCGGCCACAGGCTATCGGTAC 1684
Db
Qy 597 CGTGTGATCGGCCCGCCACCCCGTGAACATCATCGGCCGCAACATGTGACCCAGCTGGG 656
Db
Qy 1685 CGTGTGTGGGGCGCCACACCCGTCAACATCATCGGACGCAACCTGTGTGACGAGATCGG 1744
Db
Qy 657 CTGCACCTGNACTTCCCATCAGCCCATCGAGACCGTGTGCCCTGAGCTGAGCGCGG 716
Db
Qy 1745 TTGCACGCTGAACCTTCCCATTTAGCCCTATCGAGACCGTACCGGTGAAGCTGAAGCCCGG 1804
Db
Qy 717 CATGAGCGGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAAGATCAAGGCCCTGAC 776
Db
Qy 1805 GATGGAAGGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAAGATCAAGGCACTGGT 1864
Db
Qy 777 CGCCATCTGCGAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCC 836
Db
Qy 1865 GGAGATTTGCACAGAGATGGAAGAGGAGGAAATCTCCAAGATTGGGCTGAGAACCC 1924
Db
Qy 837 CTACAAACCCCGCTGTGCCCATCAAGAGAGGACAGCACCAAGTGGCGCAGCTGGT 896
Db
Qy 1925 GTACAAACCGCGGTGTTCGCAATCAAGAGAGGAACTCGACGAAATGGCGCAAGCTGGT 1984
Db
Qy 897 GGACTTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGACGCTGGGCATCCC 956
Db
Qy 1985 GGACTTCGCGAGCTGAACAGCGCACCGAAGACTTCTGGGAGGTTCACTGGGCATCCC 2044
Db
Qy 957 CCACCCCGCGCTGGAAGAGAGAGCGGTGACCGTGTGGAACGTTGGGCGCAGCCTA 1016
Db
Qy 2045 GCACCCCGCAGGGCTGAAGAGAGAAATCCGTGACCGTACTGAGTGTGGTGATGCTA 2104
Db
Qy 1017 CTTCAAGCGTCCCTGACGAGGACTTCCGCAAGTACACCGCCTTCAACATCCCCAGCAT 1076
Db
Qy 2105 CTTCTCCGTTTCCCTGGAAGAGACTTCAGGAAGTACACTGCGCTTCAATTCCTTCGAT 2164
```

```
Qy 1077 CAACAACGAGACCCCGGCATCGCTACCACTACAGTGTCTGCCCCCAGGGCTGGAAGGG 1136
Db
Qy 2165 CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTCGTGCCCCCAGGGCTGGAAGGG 2224
Db
Qy 1137 CAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGCGCTTCGCGGCCGCAA 1196
Db
Qy 2225 CTCTCCCGCAATCTTCAGAGTAGCATGACCAAAATCTCGAGCCTTTCCGCAACAGAA 2284
Db
Qy 1197 CCGCGAGATCGTGTCTTACCA-----GGCCCCCTGTACGTGGGCGACGACCTGGAGAT 1250
Db
Qy 2285 CCGCGACATCGTCATCTATCAGTACATGATGACTTGTGTCGTGGGCTCTGATCTAGAGAT 2344
Db
Qy 1251 CGGCGACGACCGCGCCAAAGATCGAGGAGCTGCGCAAGCACTGTGTGCGCTGGGCTTCAC 1310
Db
Qy 2345 AGGCGACGACCGCAACCAAGATCGAGGAGCTGCGCAGCACCTGTGTGAGGTGGGACTGAC 2404
Db
Qy 1311 CACCCCCGCAAGAGACCAAGAGGAGGCCCTTCTTCCCTTCTTGGATGGGTTACGAGCTGCA 1364
Db
Qy 2405 CACACCCGCAAGAGACCAAGAGGAGCTTCCCTTCTTGGATGGGTTACGAGCTGCA 2464
Db
Qy 1365 CCGCGACAAGTGGAGCTGCGAGCCCATCGAGCTGCCGAGAAAGAGAGCTGGAACCGTGA 1424
Db
Qy 2465 CCTGACAAATGGAACCGTGCAGCTATCGTGTGCTGCGAGAAAGACAGCTGGAATGTCAA 2524
Db
Qy 1425 CGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAA 1484
Db
Qy 2525 CGACATACAGAAGCTGTGGTGGGAAGTTGAACTGGGCGCAGTCAGATTTACCCAGGGATTAA 2584
Db
Qy 1485 GGTGCGCAGCTGTGCAAGCTGTGCGGGGCGCAAGCCCTGACCGACATCTGTGCCCCCT 1544
Db
Qy 2585 GGTGAGGAGCTGTGCAAACTCTCTCGCGGAAACCAAGGCACTCAAGAGGTGATCCCCCT 2644
Db
Qy 1545 GACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGGAGATCTCTGCGAGCCGCTGCA 1604
Db
Qy 2645 AACCGAGGAGCGCGAGCTCGAATCTGGCAGAAACCGAGAGATCTCTAAAGGAGCCGCTGCA 2704
Db
Qy 1605 CGGCGTGTACTACGACCCCGCAGCAAGGACCTGTGGTGGCGGAGATCCAGAGCAGGCGCACGA 1664
Db
Qy 2705 CGGCGTGTACTATGACCCCTCCAAAGGACCTGTGCGCGAGATCCAGAAAGCAGGCGCAAGG 2764
Db
Qy 1665 CAGTGTGACCTACAGATCTACAGGAGCCCTTCAAGAACCTGGAAGCCGCGCAAGTAGCG 1724
Db
Qy 2765 CAGTGTGACCTATCAGATTTTACAGGAGCCCTTCAAGAACCTGGAAGCCGCGCAAGTAGCG 2824
Db
Qy 1725 CAAGATCGCACCGCCGACACCAACAGCTGGAAGCAGCTGACCGAGCGCTGCAAGAGAT 1784
Db
Qy 2825 CCGATGAGGGGTGCCCACTTACGACGCTCAAGCAGCTGACCGAGGCGCTGCAAGAGAT 2884
Db
Qy 1785 CGCCATGGAGAGCATCGTGTCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAA 1844
Db
Qy 2885 CACCACCGAAAGCATCGTGTCTGGGGAAAGACTCTCTAAGTTCAAGTGTCCCATCCAGAA 2944
Db
Qy 1845 GAGAGCTGGGAGACCTGTGTGAGCAGCTACTTGGCAGGCGCACTGGATCCCGGAGTGGGA 1904
Db
Qy 2945 GGAACCTGGGAAACCTGTGTGAGCAGAGTATGGCAGGCGCACTGGATTCCTGAGTGGGA 3004
Db
Qy 1905 GTTGTGAAACACCCCGCTTGTGGAAGCTGTGTACAGCTGGAAGAGGAGCCCATCAT 1964
Db
Qy 3005 GTTGTGTCAACACCCCTTCCCTTGTGTGAGCTGTGTGTACAGCTGGAAGAGGAGCCATAGT 3064
Db
Qy 1965 CGGCGCGAGACCTTCTACGTGGAAGCGCGCGCCCAACCGCGAGAGCAAGATCGGCAAGGC 2024
Db
Qy 3065 GGGCGCGCAACCTTCTACGTGGAAGCGCGCGCTTAAACAGGAGACTAAGCTGGGCAAGGC 3124
Db
Qy 2025 CGGCTACGTACCGACCGGCGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAACCA 2084
Db
Qy 3125 CGGATACGTCACTAAACCGGGGCGAGACAGAAAGTTGTCACTCTCACTGACACCAACCA 3184
Db
Qy 2085 GAGACCGAGCTGAGGCGCATCCAGCTGGCCCTTGCAGGAGCGGCGAGCGAGGTGAACAT 2144
Db
Qy 3185 GAAGACTGAGCTGAGGCGCATTTACCTCGCTTTTGCAGGACTCGGGGCTTGGAGGTGAACAT 3244
```

```
QY 2145 CGTGACCGACGACGAGTCCCTGGGCATCATCCAGGCCAGCCGCGACAGCGAGAG 2204
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3245 CGTGACAGACTCTCAGTATGCCCTGGGCATCATCCAGGCCAGCCAGAGTGAGTGC 3304
QY 2205 CGAGCTGGTGAACACAGATCATCGACGAGCTGATCAAGAAGGAGAGGTGTACTGAGCTG 2264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3305 CGAGCTGGTCAATCAGATCATCGACGAGCTGATCAAGAAGGAGAGGTCTATCTGGCCTG 3364
QY 2265 GGTGCCCGCCCAAGAGGATCATCGCGGCNAACGAGCAGATCGACAAGCTGGTGAGCAAGG 2324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3365 GGTACCCCGCCCAAGAGGCAATTGGCGGCAATGAGCAGGTGCAAGAGCTGGTCTCGGCTGG 3424
QY 2325 CATCCGCAAGGTCTGTTCTCGGAGCGGATCGA 2357
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3425 CATCGGAAGGTGCTATTCTCGATGGATCGA 3457

RESULT 9
US-09-936-572-14
; Sequence 14, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 4327
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSVNGP4-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with 20bp of the leader
; OTHER INFORMATION: sequence of HIV-1
US-09-936-572-14

Query Match 65.1%; Score 1599.4; DB 4; Length 4327;
Best Local Similarity 81.5%; Pred. No. 9e-264;
Matches 1934; Conservative 0; Mismatches 411; Indels 28; Gaps 6;

QY 12 CATGCCGAGGCGCATGAGCCAGG---CCACGAGCGCCACATCCTGTATGCGAGCGCAGCAA 68
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1106 CTTGAGGCGGCGCATGAGCCAGGTCACCACTCCGCTACCATCATGATGCGAGCGGCGCAA 1165
QY 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGCAAGGAGGGGCCACATCGC 128
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1166 CTTTCGGAACCAACGCAAGATGTCAGTGTCTCAACTGTGCAAGAGGGCCACACAGC 1225
QY 129 CCGCAACTGCGCGCCCCCGCCCAAGAGGGCTGCTGGAGTGCAGGCGCAAGAGGGCCACA 188
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1226 CCGCAACTGCGAGGGCCCCCTAGGAAAAGGGCTGTTGGMAATGTGGAAAGGAAGGACACCA 1285
QY 189 GATGAAGNACTGCACCGAGCGCCAGCCCACTTCTCCGAGGAGGACCTGGCCCTCCCCCA 248
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1286 AATGAAGATTTACTGAGAGACAGGCTAA-TTTTTTAGGAAGATCTGGCCCTCCCA 1344
QY 249 GGGCAAGGCCCGGAGTCTCCCGAGGAGCAAGAACCGCGCAACAGCCCCCACCAGCGCGA 308
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1345 AGGAAGGCCAGGGAATTTCTTCAGAGCAGACGAGCCACAGCCCAACAGAGAGA 1404
QY 309 GTTCAGAGTGGCGGG-----CGACAAACCCCGCAGCAGAGCGCGCCGAGCGCGAGGG 362
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1405 GCTTCAGGTTTGGGAAGAGACAACTCCCTCTCAGAAAGCAGGAGCGGATAGACAAGG 1464
QY 363 CA-----CCCTGAATCTCCCCAGATCATCCCTGTGGCAGCGCCCTCGTGAGCATCAA 416
```

```
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1455 AACTGTATCTTTCAGTTCCTCAGTCACTCTTTGGCAGCACCCCTCGTCAATAAA 1524
QY 417 GGTGGCGGCGCAGATCAAGAGAGGCGCTCTGTGGAACAACCGGCGCGAGCAGACCCGTGTGGA 476
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1525 GATAGGGGGGAGCTCAAGAGAGGCTCTCTGTGGAACAACCGGAGCAGACACCCGTGTGGA 1584
QY 477 GGAGNTAGAGCTGCCCCGCAAGTGGAGCGCAAGATGATCGGCGCATCGGGCGCTTCAT 536
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1585 GGAGATGTCGTTGGCAGGCGCTGGAGCCGAAGATGATCGGGGGAATCGGGGGTTTCAT 1644
QY 537 CAAGGTGCGCAGTACGACGACAGATCCTCATCGAGATCTGCGGCAAGAGAGGCGCATCGGCAC 596
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1645 CAAGGTGCGCAGTATGACGAGATCCTCATCGAATCTGCGGCCACAGGCTATCGGTAC 1704
QY 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGTGGG 656
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1705 CGTGTGCGGCGCCACACACCGTCAACATCATCGGACGCAACCTGTTGACGAGATCGG 1764
QY 657 CTGACCCCTGAACCTTCCCATCAGCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGG 716
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1765 TTGCAACGCTGAACCTTCCCATTTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 1824
QY 717 CATGACGCGCCCAAGGTGAAGCAGTGGCCCTCTGACCGAGGAGAGATCAAGGCCCTGTAC 776
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1825 GATGACGCGCCGCAAGGTCAAGCATGCGCATTTGACAGAGGAGAGATCAAGGCACTGGT 1884
QY 777 CGCCATCTGCGAGGAGATGGAGAGAGGGGCAAGATCAACCAAGATCGGCGCCGAGAACCC 836
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1885 GGAGATTTGCAAGAGATGGAAAGAGAGGAGGAAATCTCCAAGATTGGGCGCTGAGAACCC 1944
QY 837 CTACAACACCCCGTGTTCGCCATCAAGAAGAGGACAGCACCAGTGGCGCGCAAGCTGGT 896
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1945 GTACAACAACGCGCGGTTCGCAATCAAGAAGAGGACTCGACGAATGGCGCAAGCTGGT 2004
QY 897 GGACTTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCGAGCTGGGCATCCC 956
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2005 GGACTTTCGCGAGCTGAACAAGCGCACCGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2064
QY 957 GCACCCCGCGGCTGTAAGAAGAGAGCGTGTACCGTGTGAGAGCTGGGGCGAGCGCTTA 1016
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2065 GCACCCCGCGGCTGTAAGAAGAGAGAAATCCGTGACCGGTACTGTGATGTGGGTGATGCCTA 2124
QY 1017 CTTTACGCTGCGCTGGAGAGGACTTCGCGAAGTACACCGCTTTCACCATCCCAGCAT 1076
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2125 CTTTCCGCTGCGCTGGAGCGAAGACTTCAGGAAGTACACTGCTTTCACAACTCCCTTCGAT 2184
QY 1077 CAACAACGAGACCCCGGCATCCGCTACAGTACAAACGCTGTGCTGCCCGCAGGGCTGGAAGG 1136
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2185 CAACAACGAGACACCGGGGATTCGATATCAGTACAAAGTGTGCTGCCCGCAGGGCTGGAAGG 2244
QY 1137 CAGCCCCAGCATTTTCCAGAGCAGCATGACCAAGATCTCTGAGAGCGCTTCCCGCGCCGCAA 1196
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2245 CTTCCCGCAATTTTCCAGAGTAGCATGACCAAAATCTCTGGAGCTTTCCGCAAAACAGAA 2304
QY 1197 CCGCGAGATCGTATCTACCA-----GGCCCCCTGTACTGGCGCAGCGACTGGAGAT 1250
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2305 CCGCGACATCGTATCTATCATGATGATGACTGTGTACGTGGGCTCTGTATCTAGAGAT 2364
QY 1251 CGGCGCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCCTGCTGCTGGGGCTTTCAC 1310
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2365 AGGGCAGCACCGCAACAGATCGAGGAGCTGCGCCAGCACCTGTTGAGTGGGGAGCTGAC 2424
QY 1311 CACCCCCGCAAGAGACACCAAGAGGAGCCCCCTTCTGTGCCCAT-----CGAGCTGCA 1364
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2425 CACACCCGCAAGAGACACCAAGAGGAGCTCCCTTCTCTGGATGGGTTCAGAGCTGCA 2484
QY 1365 CCGCGCAAGTGGACGCTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACCGTGA 1424
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2485 CCGTCACAAATGGAACGCTGCAGCTTATCGTGTGCGAGAAAGACAGCTGGAGCTCAA 2544
QY 1425 CGACATCCAGAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCCGCATCAA 1484
```

Db 2545 CGACATACAGAGCTGTGGGAGAGTTGAACCTGGGCGAGTCAAGATTTATCCAGGGATTAA 2604  
Qy 1485 GGTGGCCAGCTGTGCAAGCTGTCTGCGGGCGCCCAAGGCCCTTGACCGACATCTGCGCCCT 1544  
Db 2605 GGTGAGGCGAGCTGTGCAAACTCTCTCGCGGAACCAAGGCACTTCACAGAGGTGATCCCT 2664  
Qy 1545 GACGAGAGAGCGAGCTGTGAGCTGTGGCGGAGAACCGCGAGATCTCTGCGGAGCCGCTGCA 1604  
Db 2665 AACCGAGAGCGCGAGCTCGAATCTGGCAGAAAACCGAGAGATCTCTAAAGGAGCGCGTGCA 2724  
Qy 1605 CGGCTGTACTACGACCCACCAAGGACCTGTGCGCGAGATCCAGAGCAGGCGCCCGA 1664  
Db 2725 CGGCTGTACTATGACCCCTCAAGACCTGATCGCGAGATCCAGAGCAGGCGCAAG 2784  
Qy 1665 CCAGTGGACCTACAGATCTTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGC 1724  
Db 2785 CCAGTGGACCTATCAGATTTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGC 2844  
Qy 1725 CAAGATGGCACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGGCGGTGCGAGAGAT 1784  
Db 2845 CCGATGAGGGGTGCCACACATCAAGCAGCTCAAGCAGCTGACCGAGGCGGTGCGAGAGAT 2904  
Qy 1785 CGCATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAA 1844  
Db 2905 CACACCGAAGCATCTGTGATCTGGGAAAGACTCTCAAGTTCAAGCTGCCCCATCCAGAA 2964  
Qy 1845 GGAGACCTGGGAGACCTGTGTGAGCGGACTCTGGCAGGCGCACTGGATCCCGAGTGGGA 1904  
Db 2965 GGAACTCTGGAAACCTGTGTGAGCAGATTTGGCAGGCGCACTGGATCTCTGAGTGGGA 3024  
Qy 1905 GTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTAACAGCTGAGAGAGGAGCCCATCAT 1964  
Db 3025 GTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTAACAGCTGAGAGAGGAGCCCATAGT 3084  
Qy 1965 CGGCGCGAGACCTTCTAGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2024  
Db 3085 GGGCGCGAAGACCTTCTAGTGGATGGGCGCGCTAAACAGGGAGACTAAGCTGGGCAAGC 3144  
Qy 2025 CGGCTAGCTGACCGACCGGGCGCGCAGAAAGATCGTGAAGCTGACCGAGACCAACCA 2084  
Db 3145 CGGATAGCTACTAACCGGGCGACAGAGAGTTGTACCTCTACTGACACCAACCA 3204  
Qy 2085 GAAGACGAGCTGAGGCGCATCTAGCTGGCCCTCGAGGACAGCGGACGAGGTGAACAT 2144  
Db 3205 GAAGACTGAGCTGAGGCGCATTTACCTCGCTTTGCGAGACTCGGCGCTGGAGGTGAACAT 3264  
Qy 2145 CGTGACGACAGCAGTACGCGCTGGGCGATCATCCAGGCGCGCGCGAGCGAGAG 2204  
Db 3265 CGTGACAGACTCTCAGTATGCGCTGGGCGATCATTTCAAGCGCGAGCCAGACGAGTGATC 3324  
Qy 2205 CGAGCTGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTG 2264  
Db 3325 CGAGCTGTGATCATGATCATCGAGCAGCTGATCAAGAGGAAAGGTCTATCTGGCGCTG 3384  
Qy 2265 GGTCCCGCCCAAGGCGCATCGCGGCAACGAGCAGATCGACAGCTGTGTGACAAAGGG 2324  
Db 3385 GGTACCGCCGACAAAGGCAATTTGCGGCAATGAGCAGGTGCGACAGCTGTGTCTCGGCTGG 3444  
Qy 2325 CATCGCAAGGTGTGTCTTGGAGCGGATCGA 2357  
Db 3445 CATCAGGAAGGTGCTATTTCTGGATGCGATCGA 3477

RESULT 10

US-09-936-572-13

; Sequence 13, Application US/09936572  
; Patent No. 6783981  
; GENERAL INFORMATION:  
; APPLICANT: UDEN, MARK  
; APPLICANT: MITROPHANOUS, KYRIACOS  
; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
; FILE REFERENCE: 078883/0137  
; CURRENT APPLICATION NUMBER: US/09/936.572

; CURRENT FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: GB 9906177.2  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 4353  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP3-codon  
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence from  
; OTHER INFORMATION: the major splice donor  
US-09-936-572-13

Query Match 65.1%; Score 1599.4; DB 4; Length 4353;  
Best Local Similarity 81.5%; Pred. No. 9e-264;  
Matches 1934; Conservative 0; Mismatches 411; Indels 28; Gaps 6;

Qy 12 CATGCGCGAGGCCATGAGCCAGG---CCACCAGCGCCAAACATCTCTGATGCGAGCGCAGCAA 68  
Db 1132 CTTGCTGAGGCCATGAGCCAGGTGACCACTCCGCTACCATCATGATGCGAGCGCGCAA 1191  
Qy 69 CTTCAAGGGCCCCAAGCGGCATCATCAAGTGTCTTCAACTGCGCGCAAGGAGGCGCCACATCGC 128  
Db 1192 CTTTCGGAACCAACGCAAGATCGTCAAGTGTCTTCAACTGTCGCAAGAGGCGCACACAGC 1251  
Qy 129 CCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGAAAGTGGCGCAAGAGGCGCACCA 188  
Db 1252 CCGCAACTGCGAGGCGCCCTAGGAAAAGGGCTGTGAAAATGTGAAAAGGAGACACCA 1311  
Qy 189 GATGAAGGAGCTGACACGCGCGCAGGCAACTCTTCGCGAGGACCTGGGCTTCCCGCA 248  
Db 1312 AATGAAAGATTGTACTGAGAGACAGGCTAA--TTTTTTAGGGAAGATCTTGGCTCTTCCACA 1370  
Qy 249 GGGCAAGCGCCCGGAGTTCCCGCAGCGAGCAGAAACCGCGCCCAACAGCCCGCACAGCGCGCA 308  
Db 1371 AGGGAAGCGCCAGGGAATTTTCTCAGAGCAGACAGCAGAGCCCAACAGCCCGCACAGAGAGA 1430  
Qy 309 GCTGCGAGGTGCGCGG-----CGCAACCCCGCAGCGAGGCGCGCGCGAGCGCGCAGG 362  
Db 1431 GCTTCAGGTTGGGGAAGAGACAACTCCCTCTCAGAAAGCAGGAGCGGATAGACAGG 1490  
Qy 363 CA-----CCCTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCGCTGGTGAGCATCAA 416  
Db 1491 AACTGTATCTTTAGCTTCCCTCAGATCACTCTTTTGGCAGCGACCCCTCGTCACAATAA 1550  
Qy 417 GGTGGGCGGCGCAGATCAAGAGGCGCTGTGGACACCGGCGCGCAGACACCGCTGTGGGA 476  
Db 1551 GATAGGGGGCGAGCTCAAGGAGGCTCTCTTGGACACCGGAGCAGACGACCCGCTGTGGGA 1610  
Qy 477 GGAGATGAGCTGCGCGCAAGTGGAAAGCCCAAGATGATCGGCGGCAATCGGCGGCTTCAT 536  
Db 1611 GGAGATGCTGTGTCAGGCGCGCTGGAAGCCGAGATGATCGGCGGAAATCGGCGGTTTCAT 1670  
Qy 537 CAAGGTGCGCAGTACGACAGATCTGTGATGAGATCTGCGGCAAGAGGCGCATCGGCAC 596  
Db 1671 CAAGGTGCGCAGTATGACAGATCTCTCATCGAAATCTGCGGCGCACAAAGGCTTCGCTAC 1730  
Qy 597 CGTGTCTGATCGGCGCCACCGCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGG 656  
Db 1731 CGTGTCTGTTGGGCCCCACACCGCTCAACATCATCGGACGCAACCTGTTGAGCAGATCGG 1790  
Qy 657 CTGCACCTGAATTTCCCATCAGCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGG 716  
Db 1791 TTGCACGCTGAATTTCCCATTTAGCCCTATCGAGACGCTACCGGTGAAGCTGAAGCCCGG 1850  
Qy 717 CATGGACGGCGCCAGGTGAAGCGATGCGCCCTGACCGAGGAGAGATCAAGGCCCTTAC 776  
Db 1851 GATGGACGGCGCCAGGTGAAGCAATGGCCATTTGACAGAGGAGAAAGATCAAGGCACTGGT 1910



777 CGCCATCTCGAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGCAAGACC 836  
 Db |||||  
 1911 GGAGATTGTCACAGAGATGGAAAAGGAGGAAATCTCCAGATTGGGCTTGAGAACCC 1970  
 Qy |||||  
 837 CTACAAACCCCGTGTTCGCATCAAGAAGAGACAGCAACAAGTGGCGCAAGCTGGT 896  
 Db |||||  
 1971 GTACAAACACCGCGGTGTTCCCAATCAAGAAGAGAGACTCGACGAATGGCGCAAGCTGGT 2030  
 Qy |||||  
 897 GGACTTCCGCGAGCTGMAACAGCGCACCCAGAGACTTCTGGAGGTGCAGCTGGGCATCCC 956  
 Db |||||  
 2031 GGACTTCCGCGAGCTGMAACAGCGCACCGCAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2090  
 Qy |||||  
 957 CCACCCCGCGCGCTGAAGAAAGAGAGCTGACCGTGTGCGAGCTGGGCGACGCGCTA 1016  
 Db |||||  
 2091 GCACCCCGCAGGGCTGMAAGAGAGAAATCCGTGACCGTACTGGATGTGGGTGATGCCTA 2150  
 Qy |||||  
 1017 CTTTACGCGTCCCTTGGACGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCAT 1076  
 Db |||||  
 2151 CTTTCTCCGTTCCCTTGGACGAGACTTCAAGAAAGTACACTGCGCTTCAAAATCCCTTCGAT 2210  
 Qy |||||  
 1077 CAAACAGAGACCCCGGATCCGCTACCGTACCAAGCTGTCTGCCAGGCTGGAGGG 1136  
 Db |||||  
 2211 CAAACAGAGACCCCGGATTCGATATCAGTACAAACGCTGTGCCCGCAGGGCTGGAAGG 2270  
 Qy |||||  
 1137 CAGCCCGCAGCATCTTCCAGAGCAGATGACCAAGATCTTGGAGCGCTTCCGCGCCCGCAA 1196  
 Db |||||  
 2271 CTTCTCCGCAATCTTCCAGGTAGATGACCAAAATCTTGGAGCTTTCGCAACAGNA 2330  
 Qy |||||  
 1197 CCCCAGAGATCTGTATCTACCA-----GGCCCCCTGTACGTGGGAGCGACCTGGAGAT 1250  
 Db |||||  
 2331 CCCCAGCATCTGTATCTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2390  
 Qy |||||  
 1251 CGGCAGACACCGGCGCAGATCGAGAGCTGCGCAGCAGCCTGCTCGCTGGGGCTTAC 1310  
 Db |||||  
 2391 AGGGCAGACCCGACCAAGATCGAGAGCTGCGCCAGCAGCCTGTTGAGGTGGGAGTAC 2450  
 Qy |||||  
 1311 CACCCCGCAAGAGACGACCAAGAGCGCCCTTCTGCGCCAT-----CGAGCTGCA 1364  
 Db |||||  
 2451 CACACCCGACAGAGACGACCAAGAGCGCTTCTTCTGATGGGTAGAGCTGCA 2510  
 Qy |||||  
 1365 CCCCAGCAAGTGGACCGTGCAGCCCATCGAGTGTCCGAGAAAGGAGAGTGGACCGTGA 1424  
 Db |||||  
 2511 CCGTCACAAATGGACCGTGCAGCTTATCGTGTGCGCAGAGAAAGACAGCTGGACTGTCAA 2570  
 Qy |||||  
 1425 CACATCCAGAGCTGTGGGCAAGCTGNACTGGGCCAGCAGATCTACCCCGGATCAA 1484  
 Db |||||  
 2571 CGACATACAGAGCTGTGGGAAAGTGAATCTGGGCCAGTCAAGTATTTACCCAGGATTA 2630  
 Qy |||||  
 1485 GGTGCGCAGCTGTGCAAGCTGCTGCGCGGCCCAAGGCCCTGACCGACATCGTGCCT 1544  
 Db |||||  
 2631 GGTGAGGAGCTGTGCNACTCTCGCGGAACCAAGGACTCACAGAGGTGATCCCT 2690  
 Qy |||||  
 1545 GACCGAGGAGCGAGCTGGAGCTGGCGAGAACCCGAGATCTCTCGCGAGCGCCGTGCA 1604  
 Db |||||  
 2691 AACCGAGGAGCGAGCTGCAACTGCGAGAAAACCGAGAGATCTTAAGAGGAGCCGTGCA 2750  
 Qy |||||  
 1605 CGCGCTGTACTAGACCCCGCAGCAAGCACTGTGTGGCGCGAGATCCAGAGCAGGCGACA 1664  
 Db |||||  
 2751 CGCGCTGTACTAGACCCCTCAAGGACCTGTATCGCGAGATCCAGAGCAGGCGCAAGG 2810  
 Qy |||||  
 1665 CCACTGGACCTACAGATCTACAGAGGCTTCAAGAACTTCAAGAACTTCAAGAACTTCAAG 1724  
 Db |||||  
 2811 CCACTGGACCTTACAGATTTACAGAGGCTTCAAGAACTTCAAGAACTTCAAGAACTTCAAG 2870  
 Qy |||||  
 1725 CCAAGTGGGCAACCGCCACCAACGAGCTGAAGCAGTGTGACCGAGGCGGTGCAAGAT 1784  
 Db |||||  
 2871 CCGGATGAGGGGTGCCACACTAAACGAGCTCAAGCAGCTGACCGAGGCGGTGCAAGAT 2930  
 Qy |||||  
 1785 CCGCATGGAGCATCTGTATCTGGGCAAGACCCCAAGTTCCCGCTGCCATCCAGNA 1844  
 Db |||||  
 2931 CACCAACCGAAGCATCTGTATCTGGGAAAGACTCTTAAGTTCAAGCTGCCCATCCAGNA 2990  
 Qy |||||  
 1845 GGAGACCTGGGAGACCTGTGTGGACCGACTACTTGGCAGGCGCACCTGGATCCCCGAGTGGGA 1904

2991 GGAAACCTGGGAAACCTGTGGACAGAGTATTGGCAGGCCACTTGGATTCTCTGAGTGGGA 3050  
 Qy |||||  
 1905 GTTCTGTGAACACCCCGCTGTTGTAAGTGTGTGTAAGTGTGTGTAAGTGTGTGTAAGTGTGTG 1964  
 Db |||||  
 3051 GTTCTGTGAACACCCCGCTGTTGTAAGTGTGTGTAAGTGTGTGTAAGTGTGTGTAAGTGTGTG 3110  
 Qy |||||  
 1965 CGGCGCCGAGACCTTCTACGTGGAGCGCGCGCCAAACCGCAGAGACCAAGATCGGCAAGGC 2024  
 Db |||||  
 3111 GGGCGCCGAAACCTTCTACGTGGATGGGGCCGCTTAACAGGGAGACTAAGCTGGGCAAGC 3170  
 Qy |||||  
 2025 CGGCTACGTGACCGACCGGGCGCGCAGAAAGATCTGTAGCCCTGACCGAGACCAACCA 2084  
 Db |||||  
 3171 CGGATACGTCACTAACCGGGCGAGACAGAGTGTCTACCCCTCACTGTACACCAACCA 3230  
 Qy |||||  
 2085 GAAGACCGAGCTGAGGGCATCCAGCTGGCCCTTGAGGACAGCGCAGGAGGTGAACAT 2144  
 Db |||||  
 3231 GAAGACTGAGCTGCGAGGCCATTTACTCGCTTTGAGGACTCGGCGCTGGAGGTGAACAT 3290  
 Qy |||||  
 2145 CGTGACCGACAGCAGTACGCCCTGGGCATCATCCAGGCCCGCAGCCGCAAGAGCGAGAG 2204  
 Db |||||  
 3291 CGTGACAGCTCTAGTATGCCCTGGGCATCATTCAGCCCGCAGCCAGAGTGTGTC 3350  
 Qy |||||  
 2205 CGAGCTGTGTAAACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTG 2264  
 Db |||||  
 3351 CGAGCTGTCAATCAGATCATCGACAGCTGATCAAGAGGAGAAAGGTCTATCTGGCCTG 3410  
 Qy |||||  
 2265 GGTGCGCCCGCAGAGGCGATCGGCGCAACGAGCAGATCGCAAGCTGCTGAGCAAGGG 2324  
 Db |||||  
 3411 GGTATCCCGCCACAAAGGCAATTTGGCGCAATGAGCAGGTGCAAGCTGGTCTCGGCTGG 3470  
 Qy |||||  
 2325 CATCGCAAGGTGCTGTTCTCTGGAGCGCATCGA 2357  
 Db |||||  
 3471 CATCAGGAAGGTGCTATTCTCTGGATGGCATCGA 3503

RESULT 11

US-09-936-572-12  
 ; Sequence 12, Application US/09936572  
 ; Patent No. 6783981  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UDEN, MARK  
 ; APPLICANT: MITROPHANOUS, KYRIACOS  
 ; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
 ; FILE REFERENCE: 078883/0137  
 ; CURRENT APPLICATION NUMBER: US/09/936,572  
 ; CURRENT FILING DATE: 2001-12-11  
 ; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
 ; PRIOR FILING DATE: 2000-03-17  
 ; PRIOR APPLICATION NUMBER: GB 9906177.2  
 ; PRIOR FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 12  
 ; LENGTH: 4642  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP2-codon  
 ; OTHER INFORMATION: Optimised HIV-1 gapol with leader sequence  
 US-09-936-572-12

Query Match 65.1%; Score 1599.4; DB 4; Length 4642;  
 Best Local Similarity 81.5%; Pred. No. 9.1e-264;  
 Matches 1934; Conservative 0; Mismatches 411; Indels 28; Gaps 6;  
 Qy 12 CATGCCCGAGGCGCATGAGCCAGG---CCACAGCGCCCAACATCTCTGATGACGCGAGCAA 68  
 Db 1421 CTGCTGAGGCCATGAGCCAGGTGACCACTCCGCTACCATCATGATGACGCGGGCAA 1480  
 Qy 69 CTTCAAGGGCGCCCAAGCGCATCATCAAGTGTCTCAACTGTGGCAAGGAGGGCCACATCGC 128  
 Db 1481 CTTTCGGAACCAACGCAAGATCGTCAAGTGTCTCAACTGTGGCAAGGAGGGCACACAGC 1540

QY 129 CCGCAACTGCGCGCCCGCCGCAAGAGGCTGTGGAAGTGCAGCAAGGAGGCGCACCA 188  
 Db 1541 CCGCAACTGCGAGGCGCCCTAGGAAAGGCGCTGTGGAATGTGGAAGGAGGACACCA 1600  
 QY 189 GATGAAGGACTGCAACGAGCGCCAGCCCAACTTCTCCGCGAGGACTGGGCTTCCCCCA 248  
 Db 1601 AATGAAGAATGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGGCTTCCCA 1659  
 QY 249 GGGCAAGGCGCGGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGCCCCACAGCGCGGA 308  
 Db 1660 AGGGAAGCGCAGGGAATTTCTTCAAGAGCAGACAGAGCCACAGCCCAACAGAGAGA 1719  
 QY 309 GCTGAGTGGCGCG------CGACAACCCCGCAGCGAGGCGCGCGAGCGCCAGGG 362  
 Db 1720 GCTTCAGGTTTGGGGAAGAGACAAACAATCTCTCTCAGAAAGCAGGAGCGGATAGACAAG 1779  
 QY 363 CA-----CCCTGAATTTCCCGCAGATCACCTGTGGCAGGCGCCCTGTGTGAGATCAA 416  
 Db 1780 AACTGTATCTTTCAGTCTTCCCTCAGATCACTTCTTGGCAGGCGACCCCTCGTCACAATAA 1839  
 QY 417 GGTGGCGCGCAGATCAAGAGGCGCTCTGTGACACCGCGCGCAGACACACCGTGTGGA 476  
 Db 1840 GATAGGGGGCAGCTCAAGAGGCTCTCTTGGACACCGGAGGAGAGACACGTTGCTGGA 1899  
 QY 477 GGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATTCGGCGGCTTCAT 536  
 Db 1900 GGAGATGTCGTTGCGAGCGCGTGAAGCCGAAGATGATCGGGGGAATCGCGGGTTTCAT 1959  
 QY 537 CAAGTGGCGCAGTACGACCGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAC 596  
 Db 1960 CAAGTGGCGCAGTATGACCGATCTCTCATCGAAATCTGCGGCCCAAGGCTATCGGTAC 2019  
 QY 597 CGTGTGATCGGCGCCCGCCCGTGAACATCATCGGCGCAACATCTGTGACCCAGCTGGG 656  
 Db 2020 CGTGTGTGGGCGCCACACCGCTCAACATCATCGGACGCAACCTGTTGACGCGATCGG 2079  
 QY 657 CTGCAACCTGAACTTCCCATCAGCCCATCGAGCCGTCGAGCCGTCGTAAGCTGAAAGCCCG 716  
 Db 2080 TTGACGCTGAATTTCCCATTAGCCCTATCGAGCGGTACCGGTGAAGCTGAAGCCCG 2139  
 QY 717 CATGACGCGCCCAAGGTGAAGCTGTCGCGCTGACCGAGAGAGATCAAGGCGCTGAC 776  
 Db 2140 GATGACGCGCCCAAGGTCAAGCAATGGCCATTGACAGAGGAGAGATCAAGGCACTGGT 2199  
 QY 777 CGCCATCTGCGAGAGATGGAAGAGGAGGCGCAAGATCAACCAAGATCGGCCCGGAGACCC 836  
 Db 2200 GGAGTTTGCAGAGATGGAAGAGGAGGAAATCTCCAAGATTGGGCTTGAGAACCC 2259  
 QY 837 CTACAACACCCCGTGTTCGCCATCAAGAAAGAGCAGCAACCAAGTGGCGCAAGCTGGT 896  
 Db 2260 GTACAACACGCGGTGTTGCAATCAAGAAAGAGGACTCGAGCAAAATGGCGCAAGCTGGT 2319  
 QY 897 GGACTTCGCGAGCTGAACAGCGGACCCAGGACTTCTGGAGGTGACGTGGGATCCC 956  
 Db 2320 GGACTTCGCGAGCTGAACAGCGGACCCAGGACTTCTGGAGGTTCAGCTGGGCAATCCC 2379  
 QY 957 CCACCCCGCGCTGAAGAGAGAGAGCGTGCAGTGTGACGCTGGGCGGAGCGCTTA 1016  
 Db 2380 GCACCCCGCGGCTGAAGAGAGAGAAATCTCGTACCGTACTGAGTGTGGGTGATGCTTA 2439  
 QY 1017 CTTGAGGTGCGCTTGGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGACGAT 1076  
 Db 2440 CTTCTCGTTCCTTGGACGAGACTTTCAGGAGTACACTGCTTTCACATCCCTTCGAT 2499  
 QY 1077 CAACAGAGACCCCGGCTATCGCTACAGTACAGTCAAGTGTGCGCCAGGCTGGAAGG 1136  
 Db 2500 CAACAGAGACACCGGGGATTCGATATCAGTACAGTCAAGTGTGCGCCAGGCTGGAAGG 2559  
 QY 1137 CAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCTGTGAGGCTTCCGCGCCCGCA 1196  
 Db 2560 CTCTCCCGCAATCTTCAGAGTAGATGACCAAAATCTTGGAGCTTTTCGCGCAACAGAA 2619

QY 1197 CCCGAGATCTGTGATCTTACCA-----GGCCCCCTGTGCTGGGCGAGCGACTGTGAGAT 1250  
 Db 2620 CCCGCACTCTGTCTATCAGTACATGATGACTTGTGACGTGGGCTCTGATCTAGAGAT 2679  
 QY 1251 CGGCGAGCACCGCGCAGATCGAGGAGCTGGGAGCAGCCTGTGCGCTGGGCTTCA 1310  
 Db 2680 AGGGGAGCACCGCACCAAGATCGAGGAGCTGGCGCAGCACCTGTTGAGGTGGGAGTCA 2739  
 QY 1311 CACCCCGCACAAGAGACACCAAGAGGAGCCCGCTTCTGCCCCAT-----CGAGCTGCA 1364  
 Db 2740 CACACCCGACAAGAGCACAAGAGGAGCTTCCCTCTCTGATGGGTAGGAGCTGCA 2799  
 QY 1365 CCGCGACAAGTGGACCGTGCAGCCCATCGAGCTGCGCAGAAAGAGAGCTGGAACCGTGA 1424  
 Db 2800 CCCTGACAAATGGACCGGTGCAGCCCTATCGTGTGCGAGAAAGACAGCTGGAATGTCAA 2859  
 QY 1425 CGACATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCGCGCATCAA 1484  
 Db 2860 CGACATACAGAACTGTGGGGAAGTTGAATCTGGGCGAGTCAGATTTTACCCAGGATTA 2919  
 QY 1485 GGTGCGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCCGACATCTGTCGCCCT 1544  
 Db 2920 GGTGAGGCGCTGTGCAAACTCTCCGCGGAACCAAGGCACTCAAGAGGTGATCCCTCT 2979  
 QY 1545 GACCGAGAGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGCCCGTGA 1604  
 Db 2980 AACCGAGAGCCGAGCTCGAACTGGCAGAAAAACGAGAGATCTTAAAGGAGCCCGTGA 3039  
 QY 1605 CGGCGTGTACTACGACCCCGCAGAGGAGCTGTGGCGGAGATCCAGAGAGGAGGCGCAGA 1664  
 Db 3040 CGGCGTGTACTATGACCCCTCCAAAGGACCTGATCGCGGAGATCCAGAGAGGAGGCAAG 3099  
 QY 1665 CCAGTGAACCTTACAGATCTACAGAGGAGCCCTTCAAGAACTGGAAGCCGCAAGTACGC 1724  
 Db 3100 CCAGTGAACCTTACAGATTTTACAGAGGAGCCCTTCAAGAACTGGAAGCCGCGAGTACGC 3159  
 QY 1725 CAAGATCGCACCGCCCAACACAGAGCTGGAAGCAGCTGACCGAGGCGGTGCAAGAT 1784  
 Db 3160 CCGGATGAGGCGTCCCGACACTAAGCAGCTCAAGCAGCTGACCGAGGCGGTGCAAGAT 3219  
 QY 1785 CGCCATCGAGAGCATCTGTGATCTGGGCGAAGACCCCAAGTTCCGCTTCCCTTCCATCCAGAA 1844  
 Db 3220 CACCAACGAAAGCATCTGTGATCTGGGGAAGAGCTCTTAAAGTTCAAGCTGCGCCATCCAGAA 3279  
 QY 1845 GGAGACCTTGGGAGACCTGTGGACCGACTCTGCGAGCCACCTGGAATCCCGGAGTGGGA 1904  
 Db 3280 GGAACCTTGGAAACCTTGTGGACAGAGATTTGGCAGGCGACCTGGATTTCTGAGTGGGA 3339  
 QY 1905 GTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGATCCAGCTGGAGAGGAGCCCATCAT 1964  
 Db 3340 GTTCGTCAACACCCCTCCCTGTGTGAAGCTGTGTGATCCAGCTGGAGAGGAGCCCATAGT 3399  
 QY 1965 CGGCGCGAGACCTTCTGATCTGGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2024  
 Db 3400 GGGCGCGCAACCTTCTACGTGATGGGCGCGCTTAAAGGAGAGCTAAGCTGGGCGAAGC 3459  
 QY 2025 CGGCTGCTGACCGACCGGCGCGCAGAGATCGTGAAGCTGACCGGAGACCAACCA 2084  
 Db 3460 CGGATACGTCACTAACCGGGGCGACAGAGAGTTGTCAACCTCTGTCAGCAACCAACCA 3519  
 QY 2085 GAAGACGAGCTGCGAGGCCATCCAGCTGGCGCTCCAGAGACAGCGGCGAGGAGTGAACAT 2144  
 Db 3520 GAAGACTGAGCTGCGAGGCCATTTACCTCGCTTTGAGGAGCTCGGCGCTGGAGGTGAACAT 3579  
 QY 2145 CGTGAACGACGAGCTGCGCCCTGGGCAATCATCCAGGCCAGCCCGACAAAGAGCGAGAG 2204  
 Db 3580 CGTGACAGACTCTCAGTATGCGCTGGGCAATCATTAAGGCCCGAGCCAGACAGAGTGAATC 3639  
 QY 2205 CGAGCTGCTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTG 2264  
 Db 3640 CGAGCTGCTCAATCAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTATCTGCGCTG 3699  
 QY 2265 GGTGCGCGCCCAACAGGGGCTCGGCGGCATCGAGCAGATCGAACAGGCTGTGTGAGCAAGGG 2324

Db 3700 GGTACCGCCCAAGGCAATGGCGCAATGAGCGGTGCGAAGCTGTCTCGGCTGG 3759  
QY 2325 CATCGCAAGGTGCTGTTCTCTGGAGCGGCATCGA 2357  
Db 3760 CATCAGGAAGGTGCTATTCTCGATGGCATCGA 3792

RESULT 12  
US-09-552-950-5  
; Sequence 5, Application US/09552950  
; Patent No. 6541248  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; TITLE OF INVENTION: Anti-Viral Vectors  
; FILE REFERENCE: 674524-2004  
; CURRENT APPLICATION NUMBER: US/09/552,950  
; NUMBER OF SEQ ID NOS: 20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 9772  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP  
US-09-552-950-5

Query Match 65.1%; Score 1599.4; DB 4; Length 9772;  
Best Local Similarity 81.5%; Pred. No. 1e-263;  
Matches 1934; Conservative 0; Mismatches 411; Indels 28; Gaps 6;

QY 12 CATGCGCGAGGCGATGAGCCAGG----CCACGAGCGCAACATCTGTATGCGAGCGAGCA 68  
Db 2193 CCGTGGCTGAGGCGATGAGCGAGTGCACCACTCCGCTACCATCATGATGCGAGCGGCGCA 2252  
QY 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGCAAGGAGGGCCACATGCG 128  
Db 2253 CTTTCGGAACCAACGCAAGATCGTCAAGTCTTCAACTGTGCAAGAGGGCCACACAGC 2312  
QY 129 CGGCAACTGCGCGCGCCCCGCAAGAAGGGCTGTGGAAGTGCAGGCGCGAGGCGGCACCA 188  
Db 2313 CGGCAACTGCGAGCGCCCCCTAGGAAAAGGGCTGTTGGAAATGTGGAAGAGGACACCA 2372  
QY 189 GATGAAGGACTGACCGAGCGCGAGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCGCA 248  
Db 2373 AATGAAGATTGTACTGTAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCCTTCCACA 2431  
QY 249 GGGCAAGGCCCGGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCCCACCAGCGCGA 308  
Db 2432 AGGGAAGGCCAGGGAATTTTCTCAGAGCAGACCCAGAGCAACAGCCCCCACCAGAGAGA 2491  
QY 309 GCTGAGGTGCGCGG-----CGACAACCCCGCAGCGAGGCGCGCGCGAGCGCGCAGGG 362  
Db 2492 GCTTCAGGTTTGGGGAAGAGACAACTCCCTCTCAGAAGCAGGAGCGGATAGACAAGG 2551  
QY 363 CA-----CCCTGAATCTCCCGAGATCACTCTGTGGCAGCGCCCTCTGGTGAAGATCAA 416  
Db 2552 AACTGTATCTTTAGTCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAAATAA 2611  
QY 417 GGTGGCGCGCAGATCAAGGAGGCCCTGTGACACCGCGCGCGAGCACCGCTGCTGA 476  
Db 2612 GATAGGGGGGAGCTCAAGAGGCTCTCTTGACACCGGAGCAGAGCACCGCTGCTGA 2671  
QY 477 GGAGATGAGCTCTCCCGGCAAGTGAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCAT 536  
Db 2672 GGAGATGCTGTTGCCAGGCGCTGGAAGCCGAAGATGATCGGCGGAATCGGCGGTTTCA 2731  
QY 537 CAAAGTGGCCAGATGACGACAGATCTGTATCGAGATCTGCGGCAAGAGGCCATCGGCAC 596  
Db 2732 CAAAGTGGCCAGATGACGACAGATCTGTATCGAGATCTGCGGCCACAGGCTATCGGTAC 2791  
QY 597 CGTGCTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGG 656

Db 2792 CGTCTGTGTGGCCCCCAACCCGTCACATCATCGGACGCAACCTGTTGACGCGATCGG 2851  
QY 657 GTGCAACCTGTAACTTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCG 716  
Db 2852 TTGCACGCTGAACTTTCCCATTTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCG 2911  
QY 717 CATGAGCGGCCCGGAGGTGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCCCTGAC 776  
Db 2912 GATGACCGGCCCGGAGGTCAAGCAATGGCCATTGACAGAGGAGAGATCAAGGCACTGT 2971  
QY 777 CGCCATCTGCGAGGAGATGAGAGAGGAGGCAAGATCAACAGATCGGCCCGCGAGAACC 836  
Db 2972 GGAGATTTCACAGAGATGGAAGAGGAGGAAATCTCCAAGATTGGCCCTGAGAACCC 3031  
QY 837 CTACAACACCCCGTGTTCGCCATCAAGAAGAGACACCAAGTGGCGGCGCAAGCTGT 896  
Db 3032 GTACAACACCGCGGTGTTCGCAATCAAGAAGAGGACTCGACGAATGGCGAAGCTGT 3091  
QY 897 GGACTTTCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGGAGGTGCGAGCTGGGATCC 956  
Db 3092 GGACTTTCGCGAGCTGAAACAGCGCACCGCAAGACTTCTGGGAGGTTCAGCTGGGATCC 3151  
QY 957 CCACCCCGCGCTGAAAGAGAGAGAGGCTGACCGTGTGTGACGTGGGCGAGCCCTA 1016  
Db 3152 GCACCCCGCGGCTGAAAGAGAGAAATCGTGACCGTACTGGATGTGGGTGATGCTTA 3211  
QY 1017 TTTAGCGTGTGCGGAGGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGAT 1076  
Db 3212 CTTCTCCGTTCCCTGAGCAAGACTTTCAGGAAGTACACTGCTCTTCAACATCCCTTCGAT 3271  
QY 1077 CAAACAGAGACCCCGCATCCGCTACAGTACAACTGTGCTGCCCGAGGGGTGGAAGG 1136  
Db 3272 CAAACAGAGACCCCGGAGTTTCGATATCAGTACAACTGTGCTGCCCGAGGGGTGGAAGG 3331  
QY 1137 CAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAA 1196  
Db 3332 CTCTCCCGCAATCTTCCAGAGTACGATGACCAAAATCTTGGAGCCCTTCCCGCAACAG 3391  
QY 1197 CCGCGAGATCTGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGGACCTGAGAT 1250  
Db 3392 CCGCGACATCTGTATCTATCAGTATCATGGATGACTTGTACGTGGGCTCTGTACTGAGAT 3451  
QY 1251 CGGCGAGACCGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGCTGCGTGGGGTTCAC 1310  
Db 3452 AGGCGAGCACCGCACCAAGATCGAGAGCTGCGCCAGCACCTGTGAGGTGGGAGTAC 3511  
QY 1311 CACCCCGCACAGAGACACCAAGAGGAGCCCGCTTCTTGGCCCAT-----CGAGCTGCA 1364  
Db 3512 CACACCCGCAAGAGCACCAAGAGGAGCCCTCCCTTCTCTGATGGGTTACGAGCTGCA 3571  
QY 1365 CCGCGCAAGTGGACCGTGCAGCCCATCGAGTGCCTCCGAGAGAGAGAGCTGAGACCGTGA 1424  
Db 3572 CCGTGAACATGAGACCGTGCAGCCCTATCGTGTGTCAGAGAGACAGCTGAGTGTCAA 3631  
QY 1425 CGACATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGATCAA 1484  
Db 3632 CGACATCAGAGCTGTGGGGAAGTTGAATCTGGGCCAGTCAAGATTTACCCAGGATTA 3691  
QY 1485 GGTGGCGCAGCTGTCAAGAGCTGCTGCGGGGGCGCAAGCCCTGACCGACATCGTGGCCCT 1544  
Db 3692 GGTGAGGCGAGTGTGCAAACTCCTCCGCGGAACCAAGGCACTCAAGAGGTGATCCCC 3751  
QY 1545 GACCGAGAGGCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGGAGCGCGTGA 1604  
Db 3752 AACCGAGAGGCGGAGCTCGAACTGGCAGAAAACCGAGAGATCTTAAAGGAGCGCGTGA 3811  
QY 1605 CGGCGTGTACTACGACCCCGCAAGGACTGTGTGCGCAGATCCAGAGAGCAGGCGCACCA 1664  
Db 3812 CGGCGTGTACTATGACCCCTCCAGAGACTGATCGCCGAGATCCAGAGAGCAGGCGAAG 3871  
QY 1665 CAGGTGGAGCTTACAGATCTTACAGAGCCCTTCAAGAGCCCTTGAAGACCGGCAAGTAC 1724

Db 3872 CCAGTGGACCTATCAGATTACAGAGGCCCTTCAAGAACTTGAAGACCGGCAATACGC 3931  
Qy 1725 CAAGATGGCCACCGCCACACAAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAAAT 1784  
Db 3932 CGGATGAGGGGTGCCCCACACTAACAGCTCAAGCAGCTGACCGAGGCGGTGCAGAAAT 3991  
Qy 1785 CGCCATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCCCTGCCATCCAGAA 1844  
Db 3992 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCTTAAGTTCAAGCTGCCCATCCAGAA 4051  
Qy 1845 GGAGACCTGGGAGACTCTGTGACCGACTTACTTGGCAGGCGCACTTGATCCCGAGTGGGA 1904  
Db 4052 GGAACCTGGGAACCTGGTGACAGATATTGGCAGGCGCACTTGGATTCTGTAGTGGGA 4111  
Qy 1905 GTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCAT 1964  
Db 4112 GTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATAGT 4171  
Qy 1965 CGGCGCGAGACCTTCTAGTGGAGCGGCGCGCCCAACCGCAGAGACCAAGATCGGCAAGGC 2024  
Db 4172 GGGCGCGAAACCTTCTACGTGGATGGGGCGCGCTAACAGGGAGACTAAGCTGGGCAAGC 4231  
Qy 2025 CGGCTACGTGACCGAGCGGCGCGCAGAAAGATCGTGAGCCTGACCGAGACCAACCA 2084  
Db 4232 CGGATACGTCTAACCAGCGGGGAGACAAAGGTTGTACCCCTCACTGACACCAACCA 4291  
Qy 2085 GAAGACCGAGCTGAGGCGCATCCAGCTGGCGCCTGCAGGACAGCGCGAGAGTGAACAT 2144  
Db 4292 GAAGACTGAGCTGAGGCGCATTTACCTCGCTTTCAGAGACTCGGGCTGGAGGTGAACAT 4351  
Qy 2145 CGTGACGACAGGACGATGAGCGCTGGGATCATTCAGGCGCCAGCGCCGCAAGAGCGAGAG 2204  
Db 4352 CGTGACGACACTCTCAGTATGCGCTGGGCATCATTCAGCGCCAGCGAGAGTGAGTC 4411  
Qy 2205 CGAGCTGGTGAACAGACATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTGAGCTG 2264  
Db 4412 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTGAGCTG 4471  
Qy 2265 GGTGCGCGCCACAAAGGCGCATCGCGCGCAAGCAGCAGATCGACAGCTGGTGAGCAAGGG 2324  
Db 4472 GGTACCGCGCCACAAAGGCGCATTCGGCGCAATGAGCAGGTTCACAGCTGGTCTCGGCTGG 4531  
Qy 2325 CATCGCAAGTGTCTTCTGGAGCGGCATCGA 2357  
Db 4532 CATCAGGAAGTGCTATTCTCTGGATGGCATCGA 4564

## RESULT 13

US-09-872-733A-6  
; Sequence 6, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; TITLE OF INVENTION: SIV ENV GENES  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872,733A  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 8366  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
; OTHER INFORMATION: of the construct pcwvgagpolBnkan containing a CMV  
; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin  
; OTHER INFORMATION: resistance gene

## US-09-872-733A-6

Query Match 62.6%; Score 1539.2; DB 4; Length 8366;  
Best Local Similarity 79.9%; Pred. No. 1.8e-253;  
Matches 1896; Conservative 0; Mismatches 448; Indels 28; Gaps 6;  
Qy 14 TGGCGAGGCGCATGAGCGAGGCCACCCAGC---GCCAACATCTCTGATGAGCGCAGCAACT 70  
Db 1857 TGGCGAGGCGCATGAGCGAGGCTGACGAACCTCGGCGACCATATATGATGAGAGAGGCAACT 1916  
Qy 71 TCAAGGGCCCCCAAGGGCATCATCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGCCC 130  
Db 1917 TCCGGAACCAAGGGAAGATCGTCAAGTGTCTCAATTTGTGCAAAAGAGGGCACACCGCA 1976  
Qy 131 GCAACTGCGCGCGCCCCCGCAAGAGGCTCTCGAAGTGGCGAAGAGGGCCACACAGA 190  
Db 1977 GGAATCTCGGCGCCCCCGGGAAGAGGGCTGTTGGNAATGTGGNAAGAGGACACCAAA 2036  
Qy 191 TGAAGGACTGACCGAGCGCCAGGCCAACTTTCTCCGCGAGGAAGCTGGCCCTTCCCCCAGG 250  
Db 2037 TGAAGAGATTGTACTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCCCTTCTTACAAG 2095  
Qy 251 GCAAGGCGCGCGAGTTCCCCAGCGAGCAGAGAACCGGCCCAACAGCCCCACAGCGCGCAGC 310  
Db 2096 GGAAGGGCCAGGGAATTTTCTTCAGAGCAGACCAAGAGCCCAACAGCCCCCACGAGAAGAGC 2155  
Qy 311 TGCAGGTGCGCGG-----CGAACACCCCGCAGCGAGGCGCGCGCCAGCGCCAGGGCA 364  
Db 2156 TTCAAGTCTGGGTGAGAGACAACTCTCCCTCAGAAAGCAGGAGCCGATAGACAAAGAA 2215  
Qy 365 -----CCCTGAACTTCCCCCAGATCACCTCTGTGGCAGCGCCCTCTGTGAGCATCAAGG 418  
Db 2216 CTGTATCTTTAACTTCTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCTCAGTAAGA 2275  
Qy 419 TGGCGCGCAGATCAAGAGAGCCCTGCTGGAACACCGGCGCCGACGACCGTGTGAGAGG 478  
Db 2276 TCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACAGTATTAGAG 2335  
Qy 479 AGATGAGCTGCGCGGCAAGTGAAGGCCAAGATGATCGGGGCGATCGGGCGCTTCATCA 538  
Db 2336 AATAGATTGTCAGAGAGATGGAACCAAAATGATAGGGGGGATCGGGGGCTTCATCA 2395  
Qy 539 AGTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAAAGCCATCGGCACCG 598  
Db 2396 AGGTGAGGCGAGTACGACAGATCTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 2455  
Qy 599 TGCTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCT 658  
Db 2456 TATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTGTGACCCAGATCGGCT 2515  
Qy 659 GCACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCA 718  
Db 2516 GCACCTTGAATCTCCCATCAGCCCTATTGAGACGGTGCCTCGTGAAGTGAAGCGGGGA 2575  
Qy 719 TGAAGCGCCCCAAGGTGAAGCAGTGGCCCTCTGACCGGAGGAGAGATCAAGGCCCTTGACCG 778  
Db 2576 TGAAGCGCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAGATCAAGGCCCTTAGTCG 2635  
Qy 779 CCATCTCGAGAGAGATGGAGAGGAGGCGAAGATCACCAAGATCGGCCCGCGAGAACCCCT 838  
Db 2636 AAATCTGTACAGAGATGGAGAGGAGGAGGAGATCAGCAAGATCGGGCTTGAGAACCCCT 2695  
Qy 839 ACAACACCCCGTGTTCGCCCATCAAGAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTG 898  
Db 2696 ACAACACTCCAGTCTTCGCAATCAAGAGAGAGGAGCAGTACCAAGTGGAGAAAGCTGGTG 2755  
Qy 899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCC 958  
Db 2756 ACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCAC 2815  
Qy 959 ACCCGCGCGCTGAAGAGAGAGAGAGCGGTGACCGCTGTGGAGCTGGGCGAGCCCTACT 1018  
Db 2816 ATCCCGCTGGGTTGAAGAGAGAAAGTCAAGTGCAGTGTGGTGTGGTGTGGTGTGGTGTGGT 2875

1019 TCAGCGTGCCCTGGACGAGGACTTCGCGCAAGTTACACGCGCTTACCATTCCCGACATCA 1078  
1078 TCAGCGTGCCCTGGACGAGGACTTCGCGCAAGTTACACGCGCTTACCATTCCCGACATCA 1078  
2876 TCTCGGTTCCCTTGGACGAGGACTTCGCGCAAGTTACACGCGCTTACCATTCCCGACATCA 2935  
1079 ACAACGAGACCCCGGCACTCCGCTACCAAGTACAACTGCTGCTGCCCGAGGCTGGAGGCA 1138  
2936 ACAACGAGACCCCGGCACTCCGCTACCAAGTACAACTGCTGCTGCCCGAGGCTGGAGGAT 2995  
1139 GCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCGCGCCCGCAACC 1198  
2996 CACCAAGCCATCTTTCAAAGCAGCATGACCAAGATCTTCGAGCCCTTCGCGCAAGCAAAACC 3055  
1199 CCGAGATCGTGATCTACCA-----GGCCCCCTGTGAGTGGCGAGGACCTGGAGATCG 1252  
3056 CAGACATCGTGATCTATCATGATGAGCAGCTCTACGTAGGAAGTGAACCTGGAGATCG 3115  
1253 GCCAGCACCGCCCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGCTTCAACA 1312  
3116 GCGAGCACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGACTGACCA 3175  
1313 CCCCCGACAAAGAACACCAAGAGGAGCCCTTCTGCGCCAT-----CGAGCTGCAACC 1366  
3176 CACACAGCAAGAACCAAGAACCAAGAGGAACTTCCCTTCTGAGTGGGCTACGAACTGCATC 3235  
1367 CCGACAAGTGGACCGTGGACGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACG 1426  
3236 CTGACAAGTGGAGAGTGCAGCCCATCTGCTGCTGAGAGGAGAGCTGAGATGTGAACG 3295  
1427 ACATCCAGAAGCTGTTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGG 1486  
3296 ACATACAGAAGCTGTTGGGCAAGTGTGAATGGGCAAGCCAGATCTACCCAGGSCATCAAG 3355  
1487 TCGCCAGCTGTGCAAGCTGTGCGCGGCGCAAGCCCTGACCGACATCTGTGCCCTGGA 1546  
3356 TTAGGCAAGCTGTGCAAGCTGTGCGAGGAACTGAGGCACTGACAGAAGTGTATCCCACTGA 3415  
1547 CCGAGAGGCGCAGCTGGAGCTGGCGGAGAACCGGAGATCTGCGCGAGCCCGTGCACG 1606  
3416 CAGAGGAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCTGAAGAGCCAGTACATG 3475  
1607 GCGTGTACTACACCCCAAGCAAGGAGCTGTTGGCGGAGATCCAGAGCAGGCGCCACGACC 1666  
3476 GAGTGTACTACACCAAGCAAGGAGCTGATCGAGAGATCCAGAGCAGGCGGCAAGGCC 3535  
1667 AGTGACCTACAGATCTTACAGAGGAGCCCTTCAAGAACTGGAAGACCGGCAAGTACGCA 1726  
3536 AATGGACCTTACCAAACTTACAGGAGGAGCCCTTCAAGAACTGGAAGACCGGCAAGTACGCA 3595  
1727 AGATGGCACCGCCCAACCAAGCAAGGAGCTGAGCAGCTGACGAGCCGTTGCAAGAGATCG 1786  
3596 GGATGAGGGGTGCCACCAACAGATGTGAAGCAGCTGACAGGAGGAGTGCAGAGATCA 3655  
1787 CCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAAAG 1846  
3656 CCACAGAGAGCATCTGTGATCTGGGCAAGACTTCCCAAGTTCAAGCTGCCCATACAGAGG 3715  
1847 AGACTGGAGACCTGTGTGGAACGACTACTGGCAGCCACCTTGGATTCCTCCGAGTGGGAGT 1906  
3716 AGACATGGGAGACATGTGTGGAACGAGTACTGGCAAGCCACCTTGGATTCCTGAGTGGGAGT 3775  
1907 TCGTGAACACCCCTTCTGTTGGAAGTGTGTGATGAGCAGTGGAGAGGAGCCCATCATCG 1966  
3776 TCGTGAACACCCCTTCTGTTGGAAGTGTGTGATGAGCAGTGGAGAGGAGCCCATCATCG 3835  
1967 GCGCGGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGAGCAGATTCGCAAGGCGG 2026  
3836 GAGCAGAGACCTTCTACGTGGATGGGCAAGCCCAAGGAGAGCAGCTGGGCAAGGCGAG 3895  
2027 GCTACGTGACCGAGCGCGCGGAGAGATCGTGAAGCTTGAACCGAGACCAACCAACGCA 2086  
3896 GCTACGTGACCAACCGGAGCAGCAAGAGTGGTGAACCTGACTGACACCAACCAACGCA 3955

2087 AGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGGTGAACATCG 2146  
3956 AGACTGAGCTGCAGGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCG 4015  
2147 TGACCGACAGCCAGTACGCGCTTGGGCAATCTCCAGGCCAGCCCGACAAAGAGCGAGAGCG 2206  
4016 TGACAGACTCAGAGTACGCACTGGGCAATCTCCAGGCAACACCGACCAATCCGAGTCAAG 4075  
2207 AGCTGGTGAACAGCATCATCGAGCAGCTGATCAAGAGGAGAGGAGTGTACCTGAGCTGGG 2266  
4076 AGCTGGTGAACAGCATCATCGAGCAGCTGATCAAGAGGAGAGGAGTGTACCTGGCATGG 4135  
2267 TGCCCGCCCAAGGAGGATCGCGGCAACGAGCAGATCGAAGCTGTGAGCAAGGCA 2326  
4136 TACCAGCACACAAGGAATTCGAGGAATGAACAGTAGATAAATTAGTCAGTGTCTGGGA 4195  
2327 TCCGCAAGGCTGTCTCTGACGCGCATCGAT 2358  
4196 TCCGGAAGGCTGTCTCTCTGACGCGGATCGAT 4227

RESULT 14  
US-09-872-733A-1  
; Sequence 1, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872,733A  
; PRIORITY FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4338  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human  
; OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene  
US-09-872-733A-1

Query Match 61.2%; Score 1503.8; DB 4; Length 4338;  
Best Local Similarity 79.8%; Pred. No. 1.7e-247;  
Matches 1896; Conservative 0; Mismatches 447; Indels 34; Gaps 9;

14 TGGCGGAGGCGCATGAGCCAGGCGCCACCAGC---GCCAATCTCTGATGAGCGCAGCAACT 70  
1085 TGGCGGAGGCGCATGAGCCAGGCTGACGAATCTCGGCGACCAATATGATGAGAGAGGCAACT 1144  
71 TCAAGGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGGAGGCGCCACATCGCC 130  
1145 TCCGGAACAGCGGAGAGATCGTCAAGTCTTCAATTGTGGCAAGAGGCGCACACCGCCA 1204  
131 GCAACTGCGCGCCCCCGCCAGAGAGGCTGTGTGAAGTGTGCGGCAAGGAGGCGCCACCGA 190  
1205 GGAAGTCTCGGCGCCCCCGGAGAGGCTGTGTGAAGTGTGGAAGGAGGACCAACAA 1264  
191 TGAAGGACTGCACCGAGCGCCAGGCGCAACTTCTCCGCGAGGAGCTGTGGCTTCCCGCAGG 250  
1265 TGAAGATGTTACTGTAGAGACAGGCTAA-TTTTATAGGGAAGATCTGGCTTCTTACAG 1323  
251 GCAAGGCGCGGAGTTCCTCCAGCAGAGCAGAACCGGCGCAACAGCCGCCACCGCCGAGC 310  
1324 GGAAGGCGGAGGATTTTCTTCAGAGCAGACGAGAGCCCAACGCCCCCAGCAGAGAGC 1383  
311 TGCAGGTGCGGG-----CGAACACCCCGCAGCGCGCGGCGCGCGCCGAGCGCCAGGCA 364

1384 TTACAGTCTGGGTAGACACAACTCCCTCAGAACGAGGAGCCGATAGACAAGGAA 1443  
1385 -----CCTGTAATTCCTCCAGATCACTCTGTGGCAGCGCCCTGTGTGAGCATCAAGG 418  
1444 CTGTATCCTTTAACTTCTCAGATCACTTTTGGCAAGGACCCCTCTGTACAGTAAGGA 1503  
419 TGGGGCGGCAGATCAAGGAGCCCTGTGTGGACACCGGGCCGACACACCGTGTGTGGAGG 478  
1504 TCGGGGGGCACTCAAGGAAGCGCTGTGTGATACAGGAGCAGATGATACAGTATTAGAA 1563  
479 AGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGGGGATCGCGGCTTCATCA 538  
1564 AAATGAGTTTGCAGGAAGATGGAACCAAAATGATAGGGGGATCGGGGGCTTCATCA 1623  
539 AGGTGCGCAGTACGACAGATCTCTGATCGAGATCTGGGCAAGAGGCCATCGGCACCG 598  
1624 AGGTGAGCAGTACGACAGATCTCATAGAAATCTGTGGACATPAAACCTATAGGTACAG 1683  
599 TGCTGATCGGGCCCAACC-----CCGTGAACATCATCGGCCGCAACATCTGACCCAGCTG 654  
1684 TATTAGTAGGACCTACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATC 1743  
655 GGCTGCACTTGAATTCCTCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCC 714  
1744 GGCTGCACTTGAATTCCTCCATCAGCCCTATTGAGACGGTCCCGTGAAGTGAAGCCG 1803  
715 GGCTGAGCGGCCCAAGGTGAAGCAGTGGGCCCTGACCGAGGAGAGATCAAGGCCCTG 774  
1804 GGGATGAGCGGCCCAAGGTCAAGCAATGGCCATTGACAAAGAGAGATCAAGGCCCTTA 1863  
775 ACCGCACTCTGAGGAGATGAGAGAGAGGCGCAAGATCAACAAGATCGGCCCGAGAAC 834  
1864 GTCGAAATCTGTACAGAGATGAGAGAGAGGAGAGATCAGCAAGATCGGCCCTGAGAAC 1923  
835 CCCTACAAACCCCGCTGTCGCAATCAAGAGAGAGACACCAAGTGGCGCAAGCTG 894  
1924 CCCTACAACTCTGCACTTTCGCAATCAAGAGAGAGACATACCAAGTGGAGAAAGCTG 1983  
895 GTGACTTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCATC 954  
1984 GTGACTTTCAGAGCTGAACAAGAGACTCAGGACTTCTGGGAAGTTTCACTGGGCATC 2043  
955 CCCACCCCGCGCTGAAGAAGAAGAGCGTGCCTGTGGAGCTGGGCGACGCC 1014  
2044 CCACATCCCGCTGGTGAAGAAGAAGTCACTGACAGTGTGGATGTGGGTGATGCC 2103  
1015 TACTTCAGCTGCGCTGAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGC 1074  
2104 TACTTCTCCCTTGGACGAGACTTTCAGGAAGTACACTGCTTCAAGATACCTAGC 2163  
1075 ATCAACACGAGACCCCGGATCCGCTACCAAGTACCAAGTGTGCTGCCCGAGGCTGGAAG 1134  
2164 ATCAACACGAGACACCGGATCCGCTACCAAGTACCAAGTGTGCTGCCACAGGATGGAAG 2223  
1135 GGCAGCCCGCATCTTTCAGAGCAGATGACCAAGATCTCGAGCCCTTCCGCGCCGC 1194  
2224 GGATCAGCAGCATCTTTCAGAGCAGATGACCAAGATCTCGAGCCCTTCCGCAAGCAA 2283  
1195 AACCCTGAGTCTGTATCA-----GGCCCCCTGTACGTGGGCGAGGACCTGGAG 1248  
2284 AACCAGACATCTGTATCTATCTATCATGATGACACCTCTACGTAGGAAGTGAACCTGGAG 2343  
1249 ATC-GGGCAGACCGCGCAAGATCGAGGACTCGCAAGCACTCTGCTGCGCTGGGCTT 1307  
2344 ATCGGGCAGACAGGACCAAGATCGAGGACTGAGACAGCATCTGTTGAGTGGGACT 2403  
1308 CACACCCCGACAGGAAGCAACGAGAGGCCCTTCTGCCCCAT-----CGAGCT 1361  
2404 GACCACACGACAGGAAGCAACGAGAGACTCTCTTCTGTTGGATGGGCTACGAAT 2463  
1362 GCACCCCGCAAGTGGAGCGTGGACCCATCGAGCTGCGGAGAGAGAGACTGGACGT 1421  
2464 GCATCTTGACAAGTGGACAGTGGACCCCATCTGCTGCTGAGAGGACAGCTGGACTGT 2523

QY 1422 GAACGACATCCAGAGCTGTGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCGGCAT 1481  
Db 2524 GAACGACATCAGAAAGCTCGTGGCAAGTGTAACTGGGCAAGCCAGATCTTACCCAGCAT 2583  
QY 1482 CAAGTGGCGCAGCTGTGCAAGCTGTGCGGGCGCCAAAGCCCTGACCGCATCGTGCC 1541  
Db 2584 CAAGTGGCGCAGCTGTGCAAGCTGTGCGGGCGCCAAAGCCCTGACCGCATCGTATCC 2643  
QY 1542 CTTGACCGGAGGAGCGGAGCTGGCGAGAACCGCGAGATCTCTGCGGAGCCCGT 1601  
Db 2644 ACTGACGAGAGAGAGAGCTTAGAACTGGCAGAGAACCGAGAGATCTCTGAAAGAGCCAGT 2703  
QY 1602 GCACGGGTGTACTACGACCCAGCAAGCACTGTGTGGCGGAGATCTCAGAAAGCAGGCCA 1661  
Db 2704 ACATGGAGTGTACTACGACCCAGCAAGCACTGTATCGCAGAGATCCAGAGCAGGGCA 2763  
QY 1662 CGACAGTGGACCTTACCGAGTCTACGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTA 1721  
Db 2764 AGGCCAATGGACCTTACCAATCTTACAGGAGCCCTTCAAGAACCTTGAAGACGAGCAAGTA 2823  
QY 1722 CGCCAGATGCGACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAA 1781  
Db 2824 CGCAAGATGAGGGTGCACACCAAGATGTGAAGCAGCTGACAGAGGCGAGTGCAGAA 2883  
QY 1782 GATCGCATGAGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCA 1841  
Db 2884 GATCACCACAGAGAGCATCTGTATCTGGGCAAGACTCCCAAGTTCAGCTGCCCATACA 2943  
QY 1842 GAAGGAGACTTGGAGAGCTGTGTGAGACGACTTGTGGCAGGACCTTGGATCCCGAGTG 1901  
Db 2944 GAAGGAGACTTGGAGAGCATGTGTGAGACGAGTGTGGCAAGCCACTGGATCCCTGAGTG 3003  
QY 1902 GGAGTGTGTGAACACCCCGCTGTGTGAGCTGTGTGAGCTGTGGAGAGGAGCCCAT 1961  
Db 3004 GGAGTGTGTGAACACCCCGCTGTGTGAGCTGTGTGAGCTGTGGAGAGGAGCCCAT 3063  
QY 1962 CATCGCGCGCAGAGACTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAA 2021  
Db 3064 CGTGGAGCAGAGAGACTTCTACGTGGATGGGCGAGCCACAGGAGACCAAGCTGGGCAA 3123  
QY 2022 GCGCGCTTACGTGACCGAGCGGGCGCGGAGAGATCGTGTGAGCTGTGACCGAGACCA 2081  
Db 3124 GCGAGGCTTACGTGACCAACCGAGGAGCAGCAAGAGTGTGTGACCTGACTGACACCA 3183  
QY 2082 CCAGAACCGAGCTGAGGCGCATCTCAGCTGGCCCTGCGAGCAGCGCGAGGAGTGA 2141  
Db 3184 CCAGAACGAGCTGCAAGCCATCTACCTAGCTTGTGCAAGCAGCGGAGTGA 3243  
QY 2142 CATCGTACCGACAGCGAGCTACGCGCTGGGCGATCATCCAGGCCCGCCGACCAAGAGCGA 2201  
Db 3244 CATCGTACAGACTCAGTACG-CATGGGCGATCATCCAGCACAACAGACCATCGA 3302  
QY 2202 GAGCGAGCTGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACCTGAG 2261  
Db 3303 GTCAGAGCTGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACCTGCG 3362  
QY 2262 CTGGGTCCCGCCACAGGAGCATCGGGCGCAACGAGCAGATCGACAGCTGTGTGAGCAA 2321  
Db 3363 ATGGGTACCGACACAAAGGAATTGGAGGAAATGAACAGTATGATAAATAGTACAGTGC 3422  
QY 2322 GGGCATCCGAAAGTGTCTTCTTGGACGGCATCGAT 2358  
Db 3423 TGGATCCGGAAGTGTCTTCTTGGACGGATCGAT 3459

RESULT 15  
US-09-952-060-1  
; Sequence 1, Application US/09952060  
; Patent No. 6733993  
; GENERAL INFORMATION:  
; APPLICANT: Emini, Emilio A.  
; APPLICANT: Youil, Rima



```

; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2577
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Codon optimized DNA encoding modified wt pol
US-09-952-060-1

Query Match          52.2%; Score 1282.8; DB 4; Length 2577;
Best Local Similarity 85.6%; Pred. No. 7.3e-210; Mismatches 232; Indels 12; Gaps 2;
Matches 1454; Conservative 0;

QY 672 CCCCATCAGCCCATCGAGACCGTGCCTGAGCTGAAAGCCCGGATGAGCGGCCCAA 731
DB 15 CCCCCTCTCCCATTTGAGACTGTGCTGTGAAGCTGAAGCTGAGCTGGATGGCCCCNA 74

QY 732 GGTGAAGAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACGCCCATCTGCGAGGA 791
DB 75 GGTGAAGCAGTGGCCCTGACTGAGGAGAGATCAAGGCCCTGTGTGAATCTGCACTGA 134

QY 792 GATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCTTACAAACACCCCGT 851
DB 135 GATGAGAGAGGAGGCAAAATCTCAGAGATTCGCCCGGAGAACCTTACAAACCCCTGT 194

QY 852 GTTCGCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCT 911
DB 195 GTTTGCCATCAAGAAGAGGACTCCACCAAGTGGAGGAGCTGTGTGACTTCAGGGAGCT 254

QY 912 GAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCGCCACCCCGCGGCT 971
DB 255 GAACAAGAGACCCAGGACTTCTGGAGGTGAGCTGGGCATCCGCCACCCCGCGGCT 314

QY 972 GAAGAAGAGAGAGCGTGCACCGTGTGACGTGGCGCAGCGCTACTTCAGCGTGCCT 1031
DB 315 GAAGAAGAGAGAGTGTGACGTGTGATGTGGGGATGCTACTTCTGTGCCCCCT 374

QY 1032 GGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAGAGACCCC 1091
DB 375 GGATGAGGACTTCAGGAAGTACACTGCTTCCACCATCCCTCCATCAACAAATGAGACCCC 434

QY 1092 CGGCATCCGCTACAGTACAACTGTGCTGCCAGGGCTGGAGGGGAGGCCCGCAGATCTT 1151
DB 435 TGGCATCAGTACAGTACAACTGTGCTGCCAGGGCTGGAGGGGCTGCCCTGCCATCTT 494

QY 1152 CCAGAGCAGCATGACCAAGATCTGTGAGCCCTTTCGCGCCCGCAACCCCGAGATCGTGAT 1211
DB 495 CAGATCTCATGACCAAGATCTGTGAGCCCTTTCAGGAAGCAGAACCTTGACATTTGTAT 554

QY 1212 CTACCA-----GGCCCCCTGTACGTGGGCGAGCGACCTGTGAGATGGCGCAGCACCGGC 1265
DB 555 CTACCAAGTACATGGATGACCTGTATGTGGGCTCTGACCTGGAGATTGGGCGACGACGAG 614

```

```

QY 1266 CAAGATCGAGGAGCTGCGCAAGCAGCACTGCTGGCTGGGGCTTCCACACCCCGCAAGAA 1325
DB 615 CAAGATTGAGGAGCTGAGGCGAGCACTGCTGAGGTGGGGCTTGACACCCCTTGACAAGAA 674

QY 1326 GCACCAGAAGAGAGCCCGCTTCTGCTCCCAT-----CGAGCTGCAACCCCGCAAGTGGAC 1379
DB 675 GCACCAGAAGAGAGCCCGCTTCTGCTGATGGGCTATGAGCTGCAACCCCGCAAGTGGAC 734

QY 1380 CQTGAGCCCATCGAGCTGCGGAGAGGAGCTGGAACCGTGAACGACATCCAGAAGCT 1439
DB 735 TGTGAGCCCATTTGTGCTGCTGAGAGGACTCTCTGGAATGAGATGACATCCAGAAGCT 794

QY 1440 GGTGGGCAAGCTGAAGTGGGCGAGCAGATCTACCCCGGCGCATCAAGGTGCGGCGAGCTGG 1499
DB 795 GGTGGGCAAGCTGAAGTGGGCGCTCCCAATCTACCCCTGGCATCAAGGTGAGGCGAGCTGG 854

QY 1500 CAAAGCTGCTGGCGGCGCAAGGGCCCTGACCGACATCTGTGCCCTGACCGGAGGCGCGA 1559
DB 855 CAAAGCTGCTGAGGGGCGACCAAGGGCCCTGACTGAGGTGATCCCCCTGACTGAGGAGGCTGA 914

QY 1560 GCTGAGCTGCGCGAGAAACCGCGAGATCCTGCGGAGCCCGTGCACGGCGTGTACTACGA 1619
DB 915 GCTGAGCTGCTGAGAAACAGGAGATCTGAAAGGAGCTGTGCAATGGGGTGTACTATGA 974

QY 1620 CCCCAGCAAGGACCTGGTGGCGGAGATCCAGAGCAGGGCCACGACCACTGAGCTGACCTACCA 1679
DB 975 CCCCCTCCAAAGGACCTGATTGTGATCCAGAGCAGGGCCAGGGCCAGTGGACCTTACCA 1034

QY 1680 GATCTACCCAGGAGCCCTTCAAGAACTGAAAGACCGGCAAGTACGCCAAGATCGCACCCG 1739
DB 1035 AATCTACCCAGGAGCCCTTCAAGAACTGAAAGTGGCAAGTATGCCAGATGAGGGGGGCG 1094

QY 1740 CCACACCAACGAGTGAAGAGCTGACCGAGGCGGTGAGAGATCGCCATGAGAGAGCAT 1799
DB 1095 CCACACCAATGATGTGAAGCAGCTGACTGAGGCTGTGCAAGAGATCAACCACTGAGTCCAT 1154

QY 1800 CGTGATCTGGGGCAGACCCCGAGTTCCGCTGCGCATCCAGAGGAGACTCTGGGAGAC 1859
DB 1155 TGTGATCTGGGGCAGACCCCGAGTTCAAGTGTGCCATCCAGAGAGGAGACTCTGGGAGAC 1214

QY 1860 CTGTGTGACCGACTACTGCGCAGGCCCACTGTGATCCCGAGTGGGAGTTCTGTAACACCCC 1919
DB 1215 CTGTGTGACTGAGTACTGCGAGGCCCACTGTGATCCCTGAGTGGGAGTTGTGAAACCCC 1274

QY 1920 CCCCCTGTGTGAAGCTGTGTGTTACGACTGGAAGAGAGCCCATCATCGCGCGCGAGACTTT 1979
DB 1275 CCCCCTGTGTGAAGCTGTGTGTTACGACTGGAAGAGAGCCCATTTGTGGGGGCTGAGACTTT 1334

QY 1980 CTACCTGAGCAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCTAGCTGACCGA 2039
DB 1335 CTATGTGATGGGGCTGCCAACAGGGAGACCAAGCTGGGCAAGGGCTGGCTATGTGACCA 1394

QY 2040 CCGGGCGCGGCAAGAGATCTGTGAGCTTGACCGAGACCAACCAACAGAGAGCCGAGCTGCA 2099
DB 1395 CAGGGCGAGGCGAAGAGTGTGACCTGACTGACACCAACCAACAGAGAGCTGAGCTCCA 1454

QY 2100 GGCCATTCAGCTGTGGCCCTGCGAGGACAGGGCGAGCGAGGTGAACATCGTGAACGACCA 2159
DB 1455 GGCCATCTACCTTGGCCCTCCAGGACTCTGGGCTGGAGGTGAACATTTGACTGACTGCCA 1514

QY 2160 GTACGCGCTGGGCATCATCCAGGGCCCGCCGACAGAGCGGAGCTGGTGAACCA 2219
DB 1515 GTATGCCCTGGGCATCATCCAGGGCCCGCCGATCAGTCTGAGTCTGAGCTGGTGAACCA 1574

QY 2220 GATCATTCAGAGCAGCTGATCAAGAGGAGAGTGTACTCTGAGCTGGGTGCGGCCCGCAAA 2279
DB 1575 GATCATTCAGAGCAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGGTGCGGCCCGCAAA 1634

QY 2280 GGGCATTCGGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATCCGCAAGGTGCT 2339
DB 1635 GGGCATTCGGCGGCAATGAGCAGGTGGAAGAGCTGTGTCTGCTGGCATCATGGAAGGTGCT 1694

```

Qy 2340 GTTCCTGGACGGCATCGA 2357  
|||||||  
Db 1695 GTTCCTGGATGGCATTGA 1712

Search completed: June 2, 2005, 03:13:16  
Job time : 399.052 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 07:40:05 ; Search time 7836.86 Seconds  
(without alignments)  
11933.846 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457

Sequence: 1 gtcagccaccatggcga.....gggtacacgggtgaattc 2457

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hc.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	136.4	5.6	330	9	CL293849
2	100	4.1	2886	9	CL967755
3	90.2	3.7	2598	3	AV103647
4	88.8	3.6	3069	9	CL973991
5	86.6	3.5	1398	9	CL961989
6	85.6	3.5	951	4	BM321451
7	84.2	3.4	1941	9	CL971508
8	80.6	3.3	869	7	CK159167
9	80.2	3.3	1132	4	BM320864
10	79.6	3.2	1165	4	BM320900
11	79.6	3.2	1680	9	CL982770
12	79	3.2	867	4	BM321430
13	78.4	3.2	1725	9	CL978463
14	78.4	3.2	2031	9	CL974989
15	78.2	3.2	2697	9	CL952258
16	78.2	3.2	2853	9	CL974397
17	78	3.2	1485	9	CL970981
18	77.4	3.2	1509	9	CL959255
19	77.4	3.2	2559	9	CL982027
20	77	3.1	11691	9	CL962901
21	76.8	3.1	2682	9	CL969033
22	76.6	3.1	1550	4	BM321022
23	76.2	3.1	1401	9	CL962721
24	76.2	3.1	3249	9	CL945510

25 75 3.1 545 4 BI724851 1031075E0  
26 75 3.1 862 4 BM321023  
27 74.4 3.0 914 8 BZ568300 pac92-164  
28 74.2 3.0 2313 9 CL982362 OsIFSC047  
29 74 3.0 1290 9 CL972679  
30 74 3.0 2072 3 CR603312  
31 73.8 3.0 2433 9 AY401196  
32 73.2 3.0 853 4 BM321393  
33 72.8 3.0 566 4 BM587428  
34 72.6 3.0 788 6 CB643171  
35 72.6 3.0 849 3 CR718981  
36 72.4 2.9 753 9 CC675888  
37 72.4 2.9 2523 9 CL974879  
38 72.2 2.9 602 7 CV057146  
39 72.2 2.9 640 2 BE601575  
40 72.2 2.9 688 6 CB648640  
41 72.2 2.9 764 6 CB651670  
42 72.2 2.9 766 6 CB642928  
43 72.2 2.9 809 6 CB641397  
44 72.2 2.9 841 6 CB651502  
45 72.2 2.9 851 9 CG260623

#### ALIGNMENTS

RESULT 1  
CL293849/c  
LOCUS: CL293849 330 bp DNA linear GSS 12-FEB-2004  
DEFINITION: 02S0349-08A1-C03 UniformMu MutAIL Library Zea mays genomic clone  
ACCESSION: 02S0349-08A1-C03, genomic survey sequence.  
VERSION: CL293849  
KEYWORDS: GSS.  
SOURCE: Zea mays  
ORGANISM: Zea mays  
REFERENCE: 1 (bases 1 to 330)  
AUTHORS: Latshaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.  
TITLE: Sequence tagged transposon insertions from the UniformMu maize population  
JOURNAL: Unpublished (2003)  
COMMENT: Contact: Donald R. McCarty  
Plant Molecular and Cellular Biology Program  
University of Florida  
PO 110690 Gainesville, FL 32611-0690, USA  
Tel: 352-392-1928 x322  
Email: drmc@ufl.edu

FEATURES  
source  
1. .330  
Location/Qualifiers  
Class: transposon insertion site.  
Sequence flanking probable Mu insertion site in UniformMu line:  
02S0349-08, Primer set: A  
/mol\_type="genomic DNA"  
/strain="W22 (ACR, bz1-m9)"  
/cultivar="UniformMu"  
/db\_xref="taxon:4577"  
/clones="02S0349-08A1-C03"  
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."

#### ORIGIN

Query Match 5.6%; Score 136.4; DB 9; Length 330;  
Best Local Similarity 66.4%; Pred. No. 2.5e-14;

Matches 211; Conservative 0; Mismatches 106; Indels 1; Gaps 1;  
 QY 421 GCGGGCAGATCAAGAGGCCCTGCTGGACACCGGGCGGACACACCGTGTGGAGAG 480  
 Db |||||  
 Db 329 GGGGGGCGAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA 270  
 QY 481 ATGAGCTCCCGCGCAAGTGAAGCCCAAGATGATCGGGCGCATCGGGGGCTTCATCAAG 540  
 Db |||||  
 Db 269 ATGACTTTGACAGGAAGATGGAACCAAAAATGATAGGGGAATGGAGGTTTATCAAA 210  
 QY 541 GTGCGCAGTACGACAGATCTCGAGATCTGCGGCAAGAGGCCATCGGCACCGTG 600  
 Db |||||  
 Db 209 GTAAGACAGTATGATCAGTACCCATAGAAATCTGTGGGCAATAAGCTATAGGTACGTA 150  
 QY 601 CTGATCGGGCCACCCCGTGAAATCATCGGCGCGCAATCTGACCCAGCTGGGGTGC 660  
 Db |||||  
 Db 149 TTATAGGACCTACACCTGTCAACATAATTGGAAGAAATCTTTGACTCAGATTGG-TGC 91  
 QY 661 ACCCTGAACTTCCCATGAGCCCATCGAGACCGTGGCGTGAAGCTGAAGCCCGGCATG 720  
 Db |||||  
 Db 90 ACCTTAAATTTTCCCATAGTCTTATTTGAACTGTACAGTAAATTTAAAGCCAGGAATG 31  
 QY 721 GACGGCCCCAAGGTGAAG 738  
 Db |||||  
 Db 30 GATGCCCCAAAGTAAAG 13

RESULT 2  
 CL967755  
 LOCUS  
 DEFINITION  
 OsIPCC015718 Oryza sativa Expressed Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza sativa (indica cultivar-group)  
 Oryza sativa (indica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE  
 AUTHORS  
 Ma,L., Wang,C., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.  
 An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis  
 Unpublished (2004)  
 JOURNAL  
 COMMENT  
 Department of Bioinformatic  
 Beijing Institute of Genomics  
 Chinese Academy of Sciences, Beijing 101300, China  
 Tel: 86-10-80481559  
 Fax: 86-10-80488676  
 Email: chenchen@genomics.org.cn  
 Rice genomic sequence.  
 Class: exon-trapped.  
 FEATURES  
 source  
 1..2886  
 /organism="Oryza sativa (indica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:39946"  
 /clone\_lib="Oryza sativa Expressed Library"  
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN  
 Query Match 4.1%; Score 100; DB 9; Length 2886;  
 Best Local Similarity 42.8%; Pred. No. 9.6e-08;  
 Matches 700; Conservative 0; Mismatches 915; Indels 21; Gaps 3;  
 QY 189 GATGAAGGACTGACCGAGCGCAGCCAACTTCTTCGCGAGGACCTGGCCCTTCCCCCA 248  
 Db |||||  
 Db 219 GGTCTAGGAGCGCGCGCATCATGGCCATCTCTCGCAACCGCGGGCGAGCGCGCC 278

QY 249 GGGCAAGGCCCGCGAGTTCCCGAGCAGAGAGAACCGCGCAACAGCCCAACAGCGCGCA 308  
 Db |||||  
 Db 279 CGACTGGCAGAGACTTCGTGGCATCTGCTCTCTCTCTCATCACTCAACCATCAGTTT 338  
 QY 309 GCTGAGGTGGGC---GGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCAGCGGCAC 365  
 Db |||||  
 Db 339 CATCGAGAGAACAAATGCGGGAAACCGCGCGCGCGCTCATGGCGCGCTCGCGCCCAA 398  
 QY 366 CTTGAATTTCCCGCAGATCACCTGTGCGAGGCGCCCTGTGTGAGCATCAAGTGGCGG 425  
 Db |||||  
 Db 399 GACCAAGTGTCTAGGGACGGGAAATGGCAGGAGCAGGACGCGTCCATCTCTGTTCCCG 458  
 QY 426 CCAGATCAAGGAGGCCCTGTGGACACCGCGCGCGGACACACCGTGTGGAGGAGATGAG 485  
 Db |||||  
 Db 459 CGACATCATCAGCATCAAGCTCGGGGACATATCCCCCGCGAGCGCGCTGTCTCGAGG 518  
 QY 486 CTTGCCCGCAAGTGAAGCCCAAGATGATCGGGGCATCGGGCGCTTTTCATCAAGGTGG 545  
 Db |||||  
 Db 519 CGACCGCTCAAGGTTCGACGAGGGGCGCTCACCGCGAGTCGATGCCCGCTCAACAAGCA 578  
 QY 546 CGAGTACGACAGATCTGTGATCGAGATCTGCGGCGAAGAGGCATCGGCACCGTGTGAT 605  
 Db |||||  
 Db 579 CGCGGGCAGGGCGCTTTCTCGGGTCCACCGTCAAGCAGGGCGAGATCGAGGCGCTCGT 638  
 QY 606 CGGCCCAACCCCGTGAACATCATCGGCGCGCAACATCTGACCCAGCTGGGCTGCACCT 665  
 Db |||||  
 Db 639 CATGCCACCGCGGTGCACCTTTCTCGGCAAGCGGGCGCACCTGTGTGGAAGCAGCCNA 698  
 QY 666 GAACTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAAGCCCGCATGAGCG 725  
 Db |||||  
 Db 699 CAACATCGGCACCTTCAGCTCGTCTCACGGCCATCGGCAACTTCTGCATCATCTCCAT 758  
 QY 726 CCCCAAGGTGAAGCGAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTG 785  
 Db |||||  
 Db 759 CGCGTGGCATGATAATCGAGATCATCTGTATATCCCGATCCAGCACCGCGGTACCG 818  
 QY 786 CGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACAACAC 845  
 Db |||||  
 Db 819 CGACGGCATCGACAACTGTCTGCTCTCTGCTATCGGCGGATCCCATCGCCATCGCCAC 878  
 QY 846 CCCCGTGTGCCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCG 905  
 Db |||||  
 Db 879 CGTGTGTCGTACCATGGCCATCGGCTCCACCGGCTGTCGACGAGGAGGCGCCATCAC 938  
 QY 906 CGAGCTGAACAGCGCACCGAGACTTCTGGAGGTGACAGTGGGCATCCCCCAACCCGCG 965  
 Db |||||  
 Db 939 CAAGCGATGACCGCCATCGAGGAGATGGCGGCATGGACGCTGTCTCGACGCAAGAC 998  
 QY 966 CGGCTGAAGAAGAGAGCGTGACCGTGTGGACGTGGCGGACGCTACTTCAGCGT 1025  
 Db |||||  
 Db 999 CGGAGCTGACGCTCAACAGCTCACCGTCGACAAAGACGTGATCGAGGTGTACGGGG 1058  
 QY 1026 GCGCTGAGCAGGAGCTTCGCAAGATCACCGCTTCCATCCCAAGCATCAACAACGA 1085  
 Db |||||  
 Db 1059 GGGCTGGAAGAAGGACTCGGTGCTCTGTACGCGCGAGGGCGTCCCGCTCGAGAACA 1118  
 QY 1086 GACCCCGGCGATCCGCTACAGATCAACGCTGTGCCCGCAGGGCTGGAAGGGAGCCCCAG 1145  
 Db |||||  
 Db 1119 GGAGCCATTGACACATCATCTGGGGCATGCTCGCCGACCCCAAGAGGGCCCGCGCG 1178  
 QY 1146 CATCTTCAGAGCAGCATGACCAAGATCTTGAGGCCCTTCGCGCGCCGCAACCCCGAGAT 1205  
 Db |||||  
 Db 1179 CATGAAGAGGTCCACTTCTCTC-----CGTTCAACCCCGTGTGAGAAGGCAC 1226  
 QY 1206 CGTGATCTACAGGCCCCCTGTAGCTGGGCGAGCGACCTTGGAGATCGGCCAGCACCGCGC 1265  
 Db |||||  
 Db 1227 GSCCATCACTTACATCGACGGCAATGGGAATGGCACAGGATCAGAGGGGCGCGCGA 1286  
 QY 1266 CAAGATCGAGAGCTGCCAAGACCTCTGCTGGCTGGGGCTTCCACACCCCGCAAGAA 1325  
 Db |||||  
 Db 1287 GCAGATCATCAGCTGTGCAAGATGAGCAAGGAGCGCCGAGAAAGGTGTCACACGCTGAT 1346  
 QY 1326 GCACCAAGAAGGAGCCCCCTTCTTCCCATCGAGTGCACCCCGCAAGTGGACCGTGCA 1385

```

Db 1347 CGACAGTACCGGAGCCGGGCTCCGGTGGTGGCGGTGGTACCAAGAGGTGCCGA 1406
Qy 1386 GCCATCAGCTGCCGAGAGGAGAGCTGGACCGTGAACACGATCCAGAAAGCTGGTGGG 1445
Db 1407 GAAGAGCAAGAGAGGAGGAGGAGCCATGCGAGTTCGTGGCTCTCTGGCTGGTTCGA 1466
Qy 1446 CAAGCTGAATGGGCGACGAGATCTACCCGGGATCAAGGTGGCGCCAGCTGTGAAGCT 1505
Db 1467 CCGCGGAGGACGACAGCGGAGACCATCCGCGCGGTGCACTTCGGCGTGAACGT 1526
Qy 1506 GCTGCGCGGCGCAAGGCTTACCGACATCGTGGCCCTGACCGAGGAGGCGGAGCTGA 1565
Db 1527 GAAGATGATACCGGCGACGAGCTCGCCATCG-----GGAGAGAGCGGCGGCGGCT 1580
Qy 1566 GCTGCGCGAGAACCGCGAGATCTCTGCGGAGCGCGTGCACGCGCGTGTACTACGACCCCGAG 1625
Db 1581 CGGGATGGGACCAACATGTACCCGTGACGACGCTGTGGGCGACAGAGCGGAGAT 1640
Qy 1626 CAAGACCTGTGGCGGAGATCCAGAACGAGGCGCACGACGAGTGAGCTTACAGATCTA 1685
Db 1641 GAGCGGCTTCCAAATCGAGAGCTGATCGAGAGGCGCGCGCTTGGCGGCTGTTCCT 1700
Qy 1686 CCAGGAGCCCTTCAAGACCTGAAGACCGGCAAGTACCCAGATGCGCCACGCGCCACAC 1745
Db 1701 GGAGCAAGTACGAGATCGTGAAGCGCTCCAGGACCGGAAGCACATCTCGGCGATGAC 1760
Qy 1746 CAAGCAGCTGAAGCAGCTGACCGGCGGTGCAGAAAGATGCCATGGAGAGCATCTGTAT 1805
Db 1761 CGGCGAGCGCTGAACGAGCGCGCGCGCTGAAGAGCGCGACATCGCGCATGAC 1820
Qy 1806 CTGGGCGCAAGACCCCT 1821
Db 1821 CGAGCCACGAGCGCT 1836

```

## RESULT 3

```

AY103647 2598 bp mRNA linear HTC 16-OCT-2002
LOCUS AY103647
DEFINITION Zea mays PC0142084 mRNA sequence.
ACCESSION AY103647
VERSION AY103647.1 GI:21206725
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2598)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 2598)
Coe,E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
Location/Qualifiers
1. 2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="WaizeDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus

```

## FEATURES

```

source
1. 2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="WaizeDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus

```

## Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

## ORIGIN

```

Query Match 3.7%; Score 90.2; DB 3; Length 2598;
Best Local Similarity 42.2%; Pred. No. 5,4e-06;
Matches 814; Conservative 0; Mismatches 1098; Indels 15; Gaps 5;
Qy 296 CCACAGCGCGAGCTGCGAGTGGCGGCGCAACCCCGCAGCGAGGCCCGCGCGAGC 355
Db 51 CCACACCTCTCCACTCTCTGCGCACCGCGCGGCCCAACCAACACCGCAGCGAGCAA 110
Qy 356 GCAGAGGACCTGAACTTCCCCAGATCACTCTGTGGCAGCGCCCCCTGTGAGCATCA 415
Db 111 TGGCGGGCAACCGGCGCCATCGTGGAGAGCGACCCGCTGAATGGGGCGCGGGCGGG 170
Qy 416 AGTGGGGCGGCGAGATCAAGGAGGCCCTGTCTGGACACCGGCGCGCAGCACCGTGTGG 475
Db 171 AGCTGGCGGGAGCCACTTGAACGAGTGAAGCGCATGTGTGGCAGGCCCGGAGCCGG 230
Qy 476 AGGAGATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCGGGGCGCATCGCGCGCTTCA 535
Db 231 TGGTCAAGATCGAGGGTTTCCACCTCGCGGTGGCGAGTGGCGCGCTCGCTCCGCGCA 290
Qy 536 TCAAGGTGGCGCATGACACGAGATCTCTGATCGAGATCTGGGGCAGAGGCCATCGGCA 595
Db 291 AGGACGCTCGCGGCTCGCGCTCGAGCTCGACGAGGAGGCCCGCGCGCTCAAGGCCA 350
Qy 596 CCGTGTGATCGGCGCCACACCCCGTGAACATCATCTGGCGCGCAATCATCTGACCCAGCTGG 655
Db 351 GCACGAGTGGATCTCTGATCTGCTGCGCCACGGGCGGACATCTACGGCGTACCACGG 410
Qy 656 GCTGCACCTGTAATTTCCCATCGACCCCATCGAGACCGTGTCCCGTGAAGCTGAAGCCCG 715
Db 411 GCTTCGGCGGCGACCTCCACCGCGCAACAGGAGCGGCGCGCTCGCAGGTGAGAGTGC 470
Qy 716 GCATGGAGGGCCCCAAGGTGAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCCCTGA 775
Db 471 TCAGGSCATCTCAACGCGGGAATCTTCGGGCAACCGGCGAGCGGCGCACACGCTCGCGT 530
Qy 776 CCGCATCTCGGAGGAGATGGAGAGGAGGAGGCGCAAGATCAACCAAGATCGGCCCGAGAAC 835
Db 531 AGGTACCCCGCGCGCGATGCTGTGTGCGCATCAACACCCCTCTTCAGGGGCTATCTCCGGCA 590
Qy 836 CCTACAAACACCCCGCTGTTCGCCATCAAGAGAGAGGAGCAGCAAGTGGCGCGCAAGCTGG 895
Db 591 TCCGCTTCGAGATCTCGAGGCCATCAGAGCTGCTCAACACCGGTGTTCAGCCCTGCC 650
Qy 896 TGGACTTCGGGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCC 955
Db 651 TGGCGCTCGGGGCGACCATCACCGCGTGGGCGGAGTGTCCCGCTCTCTTACATCGCG 710
Qy 956 CCACCCCGCGCGCTGAAGAGAGAGAGCGTGCCTGAGCTGGAGTGGCGCGCGAGCT 1015
Db 711 GCCTCATCAGGGCGCGCGCCCAACCGCAGGCGCGTACCGG---TCGACGGGAGGAGGTGG 767
Qy 1016 ACTTCAGCGTGCCTTCGAGCGAGGACTTCGCGCAAGTACACCGCTTCAACCATCCCGAGCA 1075
Db 768 ACGCGCGGAGGCGTTCAGATCCCGGCGATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827
Qy 1076 TCAACAAAGAGACCCCCCGGATTCGCTACAGTACAAGTGTGCTGCGCCAGGGGTGAAGG 1135
Db 828 AGGAGGGCTCGCCATCGTCAACCGGCACGTCCGTGGGCTCCGCGCTCGCGGCCACCGTGA 887
Qy 1136 GCAGCCCGAGCATCTCCAGAGAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCA 1195
Db 888 TGTAACGAGCGCAACGTCTCGCGCGTCTGTCTGTGGAGGTCT---GTCCCGCTCTTCTGCG 944

```



```
QY 1196 ACCCCGAGATCGTGATCTACACAGGCCCCCTCTGCTGAGCGAGCGACCTGGAGATCGGCC 1255
Db 945 AGGTATGAAACGGCAAGCCCGAGGTACACGACACCTGACCCACAAAGCTGAAGCAACACC 1004
QY 1256 AGCACCGCGCCCAAGATCGAGGAGCTGCGCAACACCTGCTG---CGCTGGGGCTTTCAACA 1312
Db 1005 CGGGGTCCATCGAGGCGCGCCCATCATGAGACACATCTGGATGCGAGCTCTTTTCATGA 1064
QY 1313 CCCCAGACAGAACGACCAAGAGAGCCCTCTCTGCTGCCCATCGAGCTGCACCCCGACA 1372
Db 1065 AGCAGGCCAAGAAGTTGAACGAGCTGGACCCGCTCTGAAGCCCAAGCAGGACAGGTACG 1124
QY 1373 AGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCC 1432
Db 1125 CGCTCGGACAGTGCAGCGAGTGGCTGGGGCCCCAGATGAGAGTCACTCGCGCCGACACA 1184
QY 1433 AGAAGCTGGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGCGCC 1492
Db 1185 AGTCCATCGAGCGAGGTCAACTCGGTGAACGA---CAACCCGGTCACTGACGCTCCACC 1241
QY 1493 AGCTGTGCAAGTGTGCGCGCGCCGAGGCCCTGACGACATCTGTCGCCCTGACCGAGG 1552
Db 1242 GCGCAAGCGCGTGCAGCGCGCAACTTCAGGGCACCCCATCGCGCGTGTCCATGGACA 1301
QY 1553 AGGCGGAGCTGGAGCTGGCGGAGAACCGGAGATCTCGCGAGCCCGTGCACGCGGTGT 1612
Db 1302 AGCCCGCTTCGCCATCGCCAAACATCGGCAAGCTCATGTTCCGAGTTCCTCGAGCTCG 1361
QY 1613 ACTACGACCCCAAGACCTGTGTGGCGAGATCCAGAAGCAGGGCCACGACCAAGTGA 1672
Db 1362 TCAACGAGTTTACAAACACGGGCTCACTCCAACTTCGCGGACCGCAACCCCGACC 1421
QY 1673 CTTACAGATCTACAGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGC 1732
Db 1422 TGGACTACGGCTTCAAGGSCACCGAGATCGCCATGGCCCTCTACTGCTCCGAGCTCCAGT 1481
QY 1733 GCACCGCCACACCAACGAGCTGAGAGCTGACGAGCGCTGACAGAGCGCTGAGAGATGCGCATGG 1792
Db 1482 ACTCGGCMAACCCCATCAACACAGCTGAGAGCGGAGCAGCAGCAGCAACCGAGAGTGA 1541
QY 1793 AGAGCAT---CGTGATCTGGGCAAGACCCCAAGTTTCGCTGCGCTGCCATCCAGAGGAGA 1849
Db 1542 ACTCCTCGGCTCTGCTCGGCGAGGACCGCGAGCGATCGACATCTGAGCTCA 1601
QY 1850 CTGGGAGACTGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTGG 1909
Db 1602 TGTCTGTCACCTACATCTGTGGCGCTGTGCCAGGCGGTGACCTGGCGCACCTCGAGGAGA 1661
QY 1910 TGAACACCCCGCTGTTGAAGCTGTGTACACAGCTGGAGAGGAGCCCATCATCGGG 1969
Db 1662 ACATCAAGGCGTGGTGAAGAACACCGTGAACCGAGTGCCCAAGAGAGTGTGACATGA 1721
QY 1970 CCGAGACTTTTACGTGGAACGGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGCT 2029
Db 1722 ACCCTCGGCGAGCTCTCCAGCGCCCGCTTCAAGGAGAGAGCTGATCAGCGCCATCG 1781
QY 2030 AGTGACCGACCGGGCGGAGAGATGTGAGCTGACCGAGACCAACCAACAGAGA 2089
Db 1782 ACCGGAGGCGGTGTTCAGTACGGGAGGAGACCGCGCCAGCGGACGCTGCGCTGATGC 1841
QY 2090 CCGAGCTGAGGCGCATCCAGCTGGCCCTGACGAGACAGCGGAGCGAGGTGAACATCGTGA 2149
Db 1842 AGAAGCTGGCGCGTGTGGTGGACACCGCCCTCAGCAGCGGCGGACCGCGAGCGGAGC 1901
QY 2150 CCGACAGCCAGTACGCGCTGGGCGATCATCCAGGCCCGAGCCGACAGAGCGAGAGCGAGC 2209
Db 1902 CTTCCGTGTTCTCCAAATCAACAGGTTTCAGGAGGAGCTCCGCGGCTGCTGCCCGAGG 1961
QY 2210 TGGTGA 2216
Db 1962 AGGTGA 1968
```

```
RESULT 4
CL973991
LOCUS OsIFC025416 Oryza sativa Express Library Oryza sativa (indica
DEFINITION cultivar-group) genomic, genomic survey sequence.
ACCESSION CL973991
VERSION CL973991.1 GI:52402507
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 3069)
AUTHORS Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
Location/Qualifiers
source 1..3069
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "

ORIGIN



Query Match 3.6%; Score 88.8; DB 9; Length 3069;  
Best Local Similarity 41.5%; Pred. No. 9.8e-06;  
Matches 652; Conservative 0; Mismatches 917; Indels 3; Gaps 1;



QY 37 ACCAGCGCAACATCTGATCGAGCGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAG 96  
| | | | |  
Db 625 AGCGCATCGGACAGACGCTGCGAGCTACCTTACCGGAGAAAGATCCGGAG 684  
| | | | |  
QY 97 TGCTTCAACTCGGCAAGAGGGCCACATCGCCGCAACTCGCCGCCCCCGCAAGAAG 156  
| | | | |  
Db 685 CACTTCGCTGACAGCGTGGGTCCACGTCCCGCAAGTTCCGCTTCGCCACCTCCTC 744  
| | | | |  
QY 157 GGCTGCTGGAAGTGGGCAAGAGGGCCACAGATGAAGACTGCACGAGCGCCAGGCC 216  
| | | | |  
Db 745 CACGACGTATCCGCGAGGCCCAACGCCCTCTCCGCGCAGCAGAGCTCGGCGCGAGGC- 803  
| | | | |  
QY 217 AACTTTCTTCGCGAGGACTTGCGCTTCTCCCGAGGCAAGGCCCGGAGTTCCCGACCGAG 276  
| | | | |  
Db 804 --CGCGAGACGACGCTGCGCGCGCGGGTGGATGAGGCGGAGAGGCTCTCTAAGGG 861  
| | | | |  
QY 277 CAGAACCGCGCAACAGACCCCAACAGCCGCGAGCTGCAGGTGCGCGGAGAACACCCCGC 336  
| | | | |  
Db 862 CGCTCGCGCGCGCTCCAGGGGAAGAGGTACCTCGTGGTGTCTGACGACGCTCGGAGC 921  
| | | | |  
QY 337 AGCGAGGCGGCGCGAGCGCCAGGGCACCTGGAATTTCCCCACATCACTCTGTGGCAG 396  
| | | | |  
Db 922 AGGAGGAGTGGCGCTTCTCTCGCGCGCTCCCGCGGCTCGCGGCGAGTGCCTG 981  
| | | | |  
QY 397 GCGCCCTCTGTGAGCATCAAGTGGCGCGCAGATCAAGGAGGCGCTCTGTGACACCGGC 456  
| | | | |  
Db 982 CTCGTACACGCGAGGTCAAGATCAGCAGGAGTGGAGCAACCGTGGCGCGCGCGC 1041  
| | | | |  
QY 457 GCGGACGACACCGCTGCGGAGAGATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATC 516  
| | | | |  
Db 1042 GCGCTGTCGCGCACCGAGGAGCTCGGGAAGCTCGGGCGGAGCAGCGCGCGAGCTGTT 1101  
| | | | |


```

```
Qy 517 GCGGCGCATCGCGGCTTCATCAAGGTGGCGCCAGTACGACCGAGATCCTGATCGAGATCTGC 576
Db 1102 CGCGGAGGGTGTACGGCCACGGGAGCGGAGCGGAGAGATGGAGCGACTCAAGAGC 1161
Qy 577 GGCAGAGAGGCATCGGACCGTGTGATGGCCGCCACCCCGTGAACATCATCGCGCGC 636
Db 1162 CTCGTGGACTCCATGACGAAGGGCTCCAACTCGCGCTCAACATCGTCTGCGCGGA 1221
Qy 637 AACATGCTGACCCAGCTGGGTGACCGCTGAACCTTCCCATCAGCCCCATCGAGACCGTG 696
Db 1222 CTCCTTGTAGTCAAGAGAGAGATGATGGAGGCGGTGATCCATAGCTCGATGATACG 1281
Qy 697 CCGTGTAGCTGAAGCCCGCATGACGCGCCCGCCCAAGGTGAAGCACTGGCCCTGACCGAG 756
Db 1282 CCACCCCGACATCGCGGAGCAGACGCGCGCGGAGCGGAGCCGAAGAGCAGCGCTC 1341
Qy 757 GAGAGATCAAGCGCTGACCGCATCTGCGAGAGATGAGAGAGAGGGCAAGATCAC 816
Db 1342 GATGATAGCGAGCGGCGCGGACATCGACGGAGCAACGAAGAGACGAAGAGAGGAAG 1401
Qy 817 AAGATCGGCCCGAGAACCCCTACAACACCCCGTGTTCGCCCATCAAGAAGAGCAGC 876
Db 1402 CAGATGGCGACATCGATGGAACAGATCTGACGGTGTGATGAGACGACTCCCGACGAT 1461
Qy 877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGG 936
Db 1462 CTGAAGCGGTCTTCTCTACTTTCGCGGGTTACCGCGCAGACGCGATCTGCGCGGG 1521
Qy 937 GAGGTGACGTGGGATCCCCACCCCGCGGCTGAAGAAAGAGAGAGCGTCAACGCTG 996
Db 1522 AAGCTGTGCTCGGCTGTGGGTGCGGAGGGTTCATGAGGCGAAGAACGCGGACAGCGTG 1581
Qy 997 CTGCACTGGGCGAGCGCTACTTTCAGGTGCGCTTGGAGGAGTTCGCGAAGTACAC 1056
Db 1582 GAGGAGCAGCGCGAGGAGTGCTCAAGGAGTCTATCCCGGTGCTGTCGCTCGTC 1641
Qy 1057 GCCTTCCACATCCCGAGCATCAACAGAGACCCCGCATCGCTACCACTGACAAAGTG 1116
Db 1642 GAGAGCAGCGCGGCGGAGGGTGCGCGCTGACATCCACAGCGTCTGAC 1701
Qy 1117 GTGCCCCAGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTG 1176
Db 1702 TTCGTCCAGGCGGAGCGCGGACACAACTTCTCCAGCTGACAGCGGCGCGCGGC 1761
Qy 1177 GAGCGCTTCCGCGCGCGCAACCCCGAGATCTGATCTACAGGCGCCCTGTAGTGGGC 1236
Db 1762 CTCTCCAAACGCGCGCGCGCGCTCGCGCTCCGCAACACCTTACGACACCGACCTCGCC 1821
Qy 1237 AGCGACCTGGAGATCGGCGCAGCACCGGCCAAGATCGAGGAGTGGCGCAAGCAGCTGCTG 1296
Db 1822 GTGATGCTGAGGCGCGCAAGCTCAACACCTCTCTGCGATCTCCGAGCGCGCGCC 1881
Qy 1297 CGTGGGGCTTCAACACCCCGCAAGAGCAGCAGAGAGGAGCCCGCTTCTGCCCATC 1356
Db 1882 GCCAGCGCGCGCGCGCGGCTGGCAGAGAGCGCTCGAGCTCATCAATGGCGCGCC 1941
Qy 1357 GAGTGCACCCCGACAGTGAACCGTGCAGGCCATCGAGTGGCCGAGAGAGGAGCTGG 1416
Db 1942 CCCACCTTCTCGTCCACGCTCGAGTTCCTCCGCGTATGATGATCTCAAGGGCGTGGCG 2001
Qy 1417 ACCGTGAACGACATCCAGAGCTGTTGGCAAGCTGAACCTGGCGCAGCGAGATCTACCCC 1476
Db 2002 CTGCGCAGCGGAGTGTCTGCGGAGAGATCGGTGGTGTATCTCACTCAGGTACTG 2061
Qy 1477 GGCATCAAGGTGCGCAGCTGTGAAGCTGTGCGCGCGCGCAAGGCGCTTGAACGACATC 1536
Db 2062 GGCCTCTCGCACACCACTATGAGCACCTCCGAGCTCCGTCAAGAGCTCCGGAACCTG 2121
Qy 1537 GTGCCCCCTGACCGAGGAGCGGAGCTGAGCTGGCGGAGAAACCGCGAGATCTCTCGCGAG 1596
Db 2122 CAGACGTGAGCTGTGCGCAGCGAGCGTGTGAGCGCTGCGGTGCGGCTCTGCGCGAAC 2181
```

```
Qy 1597 CCGTGTGACGCG 1608
Db 2182 CCCTCGCTCGC 2193

RESULT 5
CL961989
LOCUS
DEFINITION
  OaIFCC006991 Oryza sativa Express Library Oryza sativa (indica
  cultivar-group) genomic, genomic survey sequence.
ACCESSION
  CL961989
VERSION
  CL961989.1 GI:52378720
KEYWORDS
  GSS.
SOURCE
  Oryza sativa (indica cultivar-group)
  Oryza sativa (indica cultivar-group)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
  1 (bases 1 to 1398)
  Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
  Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
  Wong, G. K. S., Deng, X. W. and Wang, J.
  An analysis of transcriptional regulation of the rice genome and
  its comparison to Arabidopsis
  Unpublished (2004)
  CONTACT: Chen Chen
  Department of Bioinformatic
  Beijing Institute of Genomics
  Chinese Academy of Sciences, Beijing 101300, China
  Tel: 86-10-80481559
  Fax: 86-10-80488676
  Email: chenchen@genomics.org.cn
  Rice genomic sequence.
  Class: exon-trapped.
  Location/Qualifiers
    1..1398
    /organism="Oryza sativa (indica cultivar-group)"
    /mol_type="genomic DNA"
    /db_xref="taxon:39946"
    /clone_lib="Oryza sativa Express Library"
    /note="Oryza sativa exon trapped genomic sequences"

ORIGIN
  Query Match 3.5%; Score 86.6; DB 9; Length 1398;
  Best Local Similarity 45.9%; Pred. No. 2.3e-05;
  Matches 506; Conservative 0; Mismatches 569; Indels 30; Gaps 5;

Qy 1 GTCGACGCCACCATGGCGGCGCATGAGCGAGCCAGCGCCAGCCACATCTCTGATGCGAG 60
Db 166 GTCGAGAGGCGAGCTCGCGAGGCGGACATGCTTCTCGTGAACACACCCACCAAG 225
Qy 61 GCGAGCACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTAAGTGGCGGAGGAGGCG 120
Db 226 GCGCGGCGCTCGGCGCGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282
Qy 121 CACATCGCCGCAACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 283 CGGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 342
Qy 181 GGCCACACAGATGAAGGACTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 343 GTCAGAGCCGCGGAGGCGCATCGAGAGATCTTCGACCAACACCGCGCGCGCGCGCGCG 402
Qy 241 TTCCCGCAGGCGAGGCGCGCGGAGTTCGCCAGCGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 403 TTCCAGATCTCTCTCAACCGCGGAGTTCCTCGCGCGGCGCGCGCGCGCGCGCGCGCG 462
Qy 301 AGCGCGAGCTGCGAGTGGCGCGGCGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGCG 360
Db 463 GCGCCCGA-----CCGCGTCTCATCGCGCGCGCGGAGACCGCGCGCGCGCGCGCG 513
Qy 361 GGCAACCTGAACCTTCCCGCGAGATCACCTGTGGCGCGCGCGCGCGCGCGCGCGCGCG 420
```

Db 514 GCCCTCAGGCGCTCAAGGACGCTAGACGCCGATGGTCCCGAGGAGGATC---CTC 570  
 Qy 421 GCGGCCAGATCAAGAGGCGCTGTGGACACGGCGCGACGACACCGTGTGGAGGAG 480  
 Db 571 ACCACCAACCTCTGGTTCGCGGAGCTGTCAAGCTCGCGCGCAACGCGTTCCTGGCCCCAG 630  
 Qy 481 ATGAGCTCTCCGCGCAAGTGGAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCAACAG 540  
 Db 631 AGGATCTGCTCGTCAACGCCATGTCCGCGCTCTGGAGGCCACCGGCGCGACGTCCGCC 590  
 Qy 541 GTGCGCCAGTACGACAGATCTGTATCGA---GATCTGCGGCAAGAAGCCATCGGCACC 597  
 Db 691 GAGTCTGCTACCGCTGGCGCAAGACTCCAGGATCGGCGCCCAAGTTTCTCAACGCCAGC 750  
 Qy 598 GTGCTGATCGGCGCCACCCCGTGAACATCATCTGGCGCGCAACATGCTGACCCAGCTGGGC 657  
 Db 751 GTGGGATTCGGTGGCTCTGCTTCAGAAAGACATCTCAACCTGCTG----- 798  
 Qy 658 TGCACCTGAATCTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGC 717  
 Db 799 TACATCTGGAGTGAATGGCTCTCCCGAGGTGGCAACTACTTGAAGCAGGTATCAAG 858  
 Qy 718 ATGGAAGCGCCCAAGGTGAAGAGTGGCCCTTGACCGAGGAGAAAGATCAAGGCCCTGACC 777  
 Db 859 ATCAAGGACTACAGAAAGCGGTTCTGTGAACCGGTTGCTGTCTCCATGTTCAACACC 918  
 Qy 778 GCATCTGGAGAGATGGAAGAGGAGGCAAGATCACAAGATCGGCCCGCGAGAACCC 837  
 Db 919 GTGGCGGCAAGAAGATCCCGTGTGGGCTTTCGCTTCAAGAAGGACACCGCGGACACC 978  
 Qy 838 TACAACACCCCGTGTGGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTG 897  
 Db 979 AGGAGAGCGCCCGCATCAAGCTCTGCAAGGGGCTGATCGGCGCAAGAGGCAAGGTGAGC 1038  
 Qy 898 GACTTCCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCC 957  
 Db 1039 ATCTAGACCCCGAGTACGAGGAGGACAGGTCCAGCGGCGACTCGCCATGAGCAAGTTTC 1098  
 Qy 958 CACCCCGCGGCTGAGAGAGAGAGAGCGGTGAGCGTGTGAGCGTGGGCGAGCGCTAC 1017  
 Db 1099 GATTGGGATCACCCGCTCCATCTCCAGCGGATGAGCGCCACCGGCTCAAGCAGGTGTCC 1158  
 Qy 1018 TTCAAGCGTCCCGTGGAGGACTTCCGCAAGTACACCGCTTCAACCATCCCGAGCATC 1077  
 Db 1159 GTGCGTGGAGCGCTACAGAGGCGGCGAGGAGCGGCGGCTGTGATCTCAACCGAG 1218  
 Qy 1078 AACACGAGACCCCGGCTACCGCTAC 1104  
 Db 1219 TGGGACGAGTTCAGGAGCTCGACTAC 1245

RESULT 6  
 BM321451  
 LOCUS  
 DEFINITION  
 rockefeller\_0.1211 Mastigamoeba balamuthi lambda ZAP II Library  
 Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase [EC  
 3.3.1.1], mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mastigamoeba balamuthi  
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
 1 (bases 1 to 951)  
 Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,  
 Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M., and  
 Philippe,H.  
 The analysis of 100 genes supports the grouping of three highly  
 divergent amoebae: Dictyostellum, Entamoeba, and Mastigamoeba  
 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
 21819461  
 11830664  
 CONTACT: Muller Miklos

Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockefeller.edu  
 Insert Length: 951 Std Error: 0.00  
 POLYA=No.  
 FEATURES  
 Location/Qualifiers  
 1. 951  
 /organism="Mastigamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
 /note="syn: Phreatamoeba balamuthi"  
 ORIGIN  
 Query Match 3.5%; Score 85.6; DB 4; Length 951;  
 Best Local Similarity 46.0%; Pred. No. 3.4e-05;  
 Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;  
 Qy 1722 CGCAGATGGCACCAGCCACCAACGACGTGAAGCAGCTGACCGAGGCGGTGCAGNA 1781  
 Db 219 CGCCAAAGATCCGCTGGTGTCTGTGCAACATCTTCTCGACGCGAGGACACGCGCGCCGC 278  
 Qy 1782 GATCGCCATGAGAGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTTCCCATCCA 1841  
 Db 279 CATCGCGAGCGCGGCTCTGCTTCTGCTGGAGGGCGGAGAACCTCCAGGAGTACTG 338  
 Qy 1842 GAAGGAGACCTGGGAGACCTGGTGACCGACTACTTGGCAGGCGCACCTGGATCCCGAGTG 1901  
 Db 339 GAGGTGACCTGGAAAGGCCCTGTCTTCCGCGCCCTACCAAGGCGCTCAGATCATGTCGA 398  
 Qy 1902 GGAGTTCTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGAGAGGAGGCCCAT 1961  
 Db 399 CGACGCGGTGACGCGACTCTGATGATCCACAAGGGGTTTCGCGCGGAGGACAAACCCCAA 458  
 Qy 1962 CATCGCGCCGAGACCTTCTAGCTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAA 2021  
 Db 459 GCTGCTGAGACGACGAGGGGCTCGAGAGGTCTGCTTGCCTTCAACAAGGTCTCAAGCA 518  
 Qy 2022 GCGCGGTACGTGACCGACCGGGGCGGAGAGATCGTGAGCTTGACCGAGACCAACCAA 2081  
 Db 519 GGTCCAGAAGGAGCAGCCCGGCTTCTGGCACAAGATCTCTCCCGAGATCCGCGGTGTCA 578  
 Qy 2082 CCAGAAGACCGAGCTGACGAGCCATCCAGCTGGCCCTCGAGGACGAGGCGGCGAGGTGAA 2141  
 Db 579 CGAGGAGACGACGACTGGCGTATGAGGCTGTACAGCTGACCGCGAGCGGCAAGCTGT 638  
 Qy 2142 CATCGTACCGACAGCAGTACGCTGGGATCATTCAGGCGCCAGCCGACCAAGACGA 2201  
 Db 639 GTTCCGCGCTCAACGTCAAGCAC---TCTNTTCAAGAGCAAGTTGACACATCTA 695  
 Qy 2202 GAGCGAGCTGGTGAACACGATCATCGACGCTGTATCAAGAGGAGAGGTTACTGTAG 2261  
 Db 696 GCGGTGCGCCACTCGCTCATCGACGCGATCAAGCGCGCACCGACGTGATCTCGCGCG 755  
 Qy 2262 TTGGGTGCGCCCGCCCAAGAGGCGATCGGCGGCAACGAGCAGATCGACAGCTGGTAGCAA 2321  
 Db 756 CAAAGTCTGCGCTGCTCGCGGCTCAAGGCGAGCGTGGGCAAGGGCTCGCGGAGTCTGCG 815  
 Qy 2322 GGGCATCCGAAGGTGCTTCTCGACGCGCATCATTCAGGCGCCAGCTGTGATCTACAGTA 2381  
 Db 816 CGGCGAGGCTGCGCGCTCATCTGTAACGAGATCGAAGGATCTGCGCGCTCGAGCGCTC 875  
 Qy 2382 CATGACGACCTGTACTGTTGGGCGAGCGCGCCCTAGGATCGA 2423  
 Db 876 GATGCGCGCTTCGAGGTCACACGCTCGAGGCGGGCTCGA 917  
 RESULT 7  
 CL971508  
 LOCUS  
 DEFINITION  
 OsIFCC021485 Oryza sativa Express Library Oryza sativa (indica  
 CL971508  
 linear GSS 21-SEP-2004



```
QY 1603 CACGGCGTGTACTAGCAAGCCAGCAAGGACCTGGTGGCC 1641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1780 GCCGACGAGTTCGAGGACAGATGAGGAGCTGGAGGGC 1818

RESULT 8
CK159167/c
LOCUS      869 bp      mRNA      linear      EST 05-DEC-2003
DEFINITION Triticum aestivum FGAS: TaLts5 Triticum aestivum cDNA,
mRNA sequence.
ACCESSION CK159167
VERSION    CK159167.1 GI:38985053
KEYWORDS   EST.
SOURCE     Triticum aestivum (bread wheat)
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Poideae; Triticeae; Triticum.
REFERENCE  1 (bases 1 to 869)
AUTHORS    Allard,F., Crosby,W.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,
            Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
            Link,W.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilsson,D.,
            Penniket,C., Roach,J.L. and Sarhan,F.
            Functional Genomics of Abiotic Stress In Wheat and Canola Crops
            Unpublished (2003)
            Contact: Wm L Crosby
            Bioinformatics
            University of Saskatchewan, Department of Computer Science
            1C101 Engineering Building, 57 Campus Drive, Saskatoon,
            Saskatchewan, S7N 5A9, Canada
            Tel: 306 966 1769
            Fax: 306 966 2033
            Email: fgas.ests@cs.usask.ca
            This sequence is the direct result of the Base calling software
            Phred (default parameters). It is the raw base calls. To aid in the
            identification of the high quality insert the software Lucy
            (default parameters) has been run on this sequence. Lucy identified
            the region [128,636].
            Plate: TaLts537, row: N column: 23.
FEATURES   Location/Qualifiers
            1..869
                /organism="Triticum aestivum"
                /mol_type="mRNA"
                /cultivar="wheat line PI 178383"
                /db_xref="taxon:4565"
                /lab_host="DH5 alpha"
                /clone_lib="Triticum aestivum FGAS: TaLts5"
                /note="Organ: Crown; Vector: pGEM-T; SSH (suppression
                subtractive hybridization) cDNA library from genotype
                PI178383 cold hardened at 2 C for 21 days and 49 days
                (equal amount of cDNA pooled together before subtraction,
                tester) and subtracted against genotype Norstar cold
                hardened at 2 C for 1 day (24 H) (driver). Modified Smart
                cDNA (Clontech)/priming and non-directional cloning"
ORIGIN
Query Match      3.3%; Score 80.6; DB 7; Length 869;
Best Local Similarity 44.9%; Pred. No. 0.00027;
Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

QY 580 AAGAAGGCCATCGGACCGCTGTGATCGGCCCGCCAGCCCGTGAACATCATCGCGCGCAAC 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 827 ATGGCGGGCGCCGACACAGAGGATCACCACACGAGGAGCAACACCAACCAACCAACCAAC 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 640 ATGCTGACACAGTGGCTGCACCTTGAACTTCCCATCATCGCCCATCGAGACCGTGGCC 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 CACAACGACGACAAACAGCGACACACGACCAACCAACCAACCAACCAACCAACCAAC 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 700 GTGAGCTGAAGCCCGCATGACGCGCCCGAGGTGAGCACTGCGCCCTGACCGGAG 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 ACCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 760 AAGATCAAGGCCCTGACCGCCATCTCGGAGGAGATGGAGAGGGGCAAGATCAACCAAG 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

	Query Match	3.3%; Score 80.2; DB 4; Length 1132;	
	Best Local Similarity	45.3%; Pred. No. 0.00032;	
	Matches	375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;	
Qy	139	CGGCCCCCGCAAGAGGGCTGTCTGAAGTTCGGCAAGGAGGGCCACACAGATGAAGCAC	198
Dd	79	CGCGCGAGGGCAAGACGACTACCGCGCGGCANCTGGTGATTCAGGACACAGAACAA	138
Qy	199	TGCACCGAGCGCCAGGCCAATTCTTTCCGCGAGGACCTGGCCTTCCCACCAGGCGCAAGGCC	258
Dd	139	TACAACAGCCCCAAGTACCGTTTGGTC-----GTCCGCTTCAACCAACAGGACATCGTC	192
Qy	259	CGCGAGTTCCCGAGCAGACAGAACCGCGCAACAGCCCCACAGCGCGGAGCTGCGAGTG	318
Dd	193	TGCCAGATCGCTACGCCAAGATCGACGGCGACCATCTCGCGCGCGCTACTCGCAC	252
Qy	319	CGCGGGCGAACACCCCGCAGCGAGCGCGCGCGGAGCGGCACCTGAACCTTCCCG	378
Dd	253	GAGCTCACCGCTTCGGGTGAAGTCTGGCGCTGACCAACTACCGCGCGCTTACGCGACT	312
Qy	379	CAGATCACCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTGGGCGGCAGATCAAGGAG	438
Dd	313	GGCTCTGTCTGGCCCGCGGTGTCTGAAGAAGCTCAACTCGACTTCCAAGTACGAGGT	372
Qy	439	GCCTCTGTGACACCGCGCGCGACGACACCTGTCTGTGAGGAGATGAGCCTGCGCGGAAG	498
Dd	373	GTCAAGAAGGTCAACGGCGAGGACTACAACTGTGAGGAGCTCGACGAGCGGCGCGCGT	432
Qy	499	TGGAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCAAGTGGCGGAGTAGACACAG	558
Dd	433	TTCAAAGCCCTGTCTGACGTGGCGCTGTGGTCCGACCTCGACTGGCGCGCGGTGTGCGC	492
Qy	559	ATCCTGATCGAGATCTCGCGCAAGAGGCGCATCGGCACCGTCTGATCGGCCCCACCCCC	618
Dd	493	GCCTCAAGGGCNATGCGACGGCGCGTCAAGTCCCC--CACAGGAGACCGCTTC	549
Qy	619	GTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGACACCTGAACTTCCCCATC	678
Dd	550	GTGGCTTCAACGGCGACAAAGAGAGCTCAACCGCGCGTCTCTCGCAAGTACATCTTC	609
Qy	679	AGCCCCATCGAGACCTGCGCTGAGCTGAAGCGCGGATGAGCGGCGCCCAAGTGAAG	738
Dd	610	GGCGGCCACGTGCGCCCGTACATGAAGCTCTCAAGGAGCAGGACCGCGCGCTTCGAC	669
Qy	739	CAGTGGCCCCGTACCGAGAGAGATCAAGGCGCTTGACCGCCATCTCGAGGAGATGAG	798
Dd	670	CGCCAGTTCTCGCGCTTACGCCAAGGGGTGTACCGCGGACATGCTCGAAGATCTAC	729
Qy	799	AAGGAGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTCAACACACCCCGTGTTCGCC	858
Dd	730	ACCGAGGCCACAAAGCAGATCGCGCGCGACCGCCTTGTCTCCCAAGCGCGCTCGAAG	789
Qy	859	ATCAAGAAGAGAGACAGCACCAAGTGGCGCAAGCTGTGTGGAATTCGGGAGCTGAACAG	918
Dd	790	CCGAGGGGGCCCAAGCCCAAGCACTGGGGCAGCGCAGGCTGACGTACCAAGAGCGCAAG	849
Qy	919	CGCACCCAGAGCTTCTGGGAGGTGAGCTGGGATCCCCCGCCCGCCCGCC	966
Dd	850	AACCGGTGCGCCAGAAAGGTTCGCGCTGGGCTACCCCGAGCGCCCC	897
RESULT 10			
BM320900			
LOCUS			
DEFINITION			
rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library			
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA			
sequence.			
BM320900			
BM320900.1 GI:18055306			
EST.			
Mastigamoeba balamuthi			
Mastigamoeba balamuthi			
ORGANISM			

Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.	
1 (bases 1 to 1165)	
Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,	
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and	
Philippe,H.	
The analysis of 100 genes supports the grouping of three highly	
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba	
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)	
21819461	
11830664	
Contact: Muller Miklos	
Laboratory of Biochemical Parasitology	
The Rockefeller University	
1230 York Avenue, New York, NY 10021, USA	
Email: mmuller@rockvax.rockefeller.edu	
Insert Length: 1165 Std Error: 0.00	
POLYA=No.	
Location/Qualifiers	
1..1165	
/organism="Mastigamoeba balamuthi"	
/mol_type="mRNA"	
/strain="ATCC 30984"	
/db_xref="taxon:108607"	
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"	
/notes="syn: Phreatamoeba balamuthi"	
ORIGIN	
Query Match	3.2%; Score 79.6; DB 4; Length 1165;
Best Local Similarity	45.1%; Pred. No. 0.00041;
Matches	428; Conservative 0; Mismatches 506; Indels 16; Gaps 3
Qy	90 CATCAAGTGTTCAACTGCGGCAAGAGGGGCCAATCGCCCAACTGCGCGCCCCCG 149
Dd	46 CGTCAAGAAACAAGCGCTACTTCAAGCGCTTCAGACCCAGTTCGTCGCGCGCGAGG 105
Qy	150 CAGAAGGGTGTCTGAAGTGGCGCAAGAGGGCCACAGATGAAGGATGACACCGAGG 209
Dd	106 CAAGACGAGACTACCGCGCGCCANCTGGTGATCCAGGACAAGACAAGTACAACAGGCC 165
Qy	210 CCAGGGCAACTTCTTCGCGGAGGACCTGGCTTCCCGCAGGGCAAGGGCCGGAGTTGCC 269
Dd	166 CAAGTACCGTTGTC-----GTCCGCTTCAACAACAGGAGATCTGTCTGCAGATCG 219
Qy	270 CAGCGAGCAGAAACCGCGCAACACAGCCCCACAGCGCGAGCTGCAGGTGCGCGGCAAA 329
Dd	220 CTAGCGCAAGATCGACGGCGACACATCTCGCGCGCTACTCGCAGAGCTACACCG 279
Qy	330 CCCCGAGGAGCGCGCGCGCGAGCGCGCAGGGCACCTGTAATCTTCCCGCAGATCACCT 389
Dd	280 CTTTCGGCGTCAAGCTCGGCTGACCAACTACGCGCGCTTACGCGACTTGGCTGTGTGT 339
Qy	390 GTGGCAGCGCCCCCTGGTGGAGCATCAAGGTGGCGGCGCAGATCAGAGGCGCTTGTGGA 449
Dd	340 GGCGCGCGTGTGTCTGAAGAAGCTCAACCTCTACTCAAGTACAGGGGTGTCAAGAAGT 399
Qy	450 CACCGCGCGCGACGACACCGTGTGGAGGAGATGAGCTTCCCGCGCAAGTGAAGGCCAA 509
Dd	400 CAACGGCGAGGACTACAACTCGAGGAGCTCGACGAGGGCGCGCGCTTCAAGGCC--- 457
Qy	510 GATGATCGCGGATCGCGGGCTTCAAGGTGGCGCAGTACGACCAAGATCTCTGATGGA 569
Dd	458 --TGCTCGACGTGGCGCTGTGTCCGACCTCGACTGGCGCGCGCTTTCGCGCGCTCAA 5



Db	636	CGCCGCGTACATGAAGCTCTCAAGGAGCAGGACGCCGCCCTTCGACCGCGAGTTCTC	695	QY	330	GTGCGAGCGCCCTGTGTGAGCATCAAGTGGGGCGCCAGATCAAGAGGCGCCCTGTCTGA	449
QY	750	GACCGAGGAGAGATCAAGGCGCTGACCGCCATCTCGAGGAGATGGAGAAGGAGGCA	809	Db	249	GAAGAAGACCAACAATCTTCAACGGGGCGCATCTTACAGTCCGTGCTGGAGAAGGAGGG	308
Db	696	GCGCTACGCGCAAGGAGGGTGTACCGCCGACATGCTCGAGAAGATCTTACACCGGCGCCA	755	QY	450	CACCGGCGCCGACGACACCGTGTCTGGAGGAGATGAGCCCTGCCCCGCAAGTGGAAAGCCCAA	509
QY	810	GATCACCAAGATCGGCGCCGAGAACCCCTTACAACACCCCGTGTTCGCCATCAAGAAGAA	869	Db	309	CGCGGCGACTACCTTCGGCAAGACGGTGCAGGTGATCCCGCAGCTCACCAACAGAGATCCA	368
Db	756	-----CAAGCAGATCCGCGCCGACCCGACCTTGTCCCAAGCGGCTCGAAGCCGA	809	QY	510	GATGATCGGC---GGCATCGGCGGCTTTCATCAAGTGGGCCAGTACGACCAGATCCTGAT	566
QY	870	GGACGACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAAACAAGCGCACCCAGGA	929	Db	369	GGATATCTCAAGCGCGCGCGGCATCGGCACCGACGCGGTGGGACGTGGCCATCGT	428
Db	810	GGCGCCCAAGCCCAAGCATGCGGCAAGCGCAGGCTGACGTACACGAGCGCAAGACCG	869	QY	567	CGAGATCTCGCGCAAGAGGCGCATCGGCACCGTGTGATCGGCCCCACCCCGCTGAACAT	626
QY	930	CTTCTGGGAGTGCAGCTGGGATCCCCACCCCGCGGCTTGAAGAGAGAGAGCGT	989	Db	429	CGAGATCGCGGACCGTGGGGACATCGAATCGTTCGGTTCCTCGAAGCCGCTGGCCA	488
Db	870	CGTCCGCCCAAGAAGAGTCCGCTGGGTACCCCGACGCCCCCAAGAGCCAGTAAATTCC	929	QY	627	CATCGGCGCGCAACATGTGTGACCCAGCTGGGGTGCACCTGTGAATTTCCCCATCAGGCCCAT	686
QY	990	GACCGTGTGAGCTGGGCGAGCGCTTCTTACGCGTGGCCCTTGGACGAGG	1039	Db	489	GATGAGCCTGCGCATGGGCGCGCAATTCGCGCTTCGTGCACCTCACCTACCTGCGCTA	548
Db	930	CGGTGCCCTGTACACGCCCATCTCCGCTGCGGGGTGGCTGTCTGCTGCTG	979	QY	687	CGAGACCGTGGCGGTGAAGCTGAAGCCCGGCATGGACGGCCCAAGGTGAAGCAGTGGCC	746
RESULT 11							
CL982770							
LOCUS							
DEFINITION	CL982770 1680 bp DNA linear GSS 21-SEP-2004						
ACCESSION	OsIFSC049024 Oryza sativa Express Library Oryza sativa (indica						
VERSION	CL982770						
KEYWORDS	GSS.						
SOURCE	Oryza sativa (indica cultivar-group)						
ORGANISM	Oryza sativa (indica cultivar-group)						
REFERENCE	1 (bases 1 to 1680)						
AUTHORS	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,						
TITLE	Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,						
JOURNAL	Wong, G. K. S., Deng, X. W. and Wang, J.						
COMMENT	An analysis of transcriptional regulation of the rice genome and						
	its comparison to Arabidopsis						
	Unpublished (2004)						
	Contact: Chen Chen						
	Department of Bioinformatic						
	Beijing Institute of Genomics						
	Chinese Academy of Sciences, Beijing 101300, China						
	Tel: 86-10-80481559						
	Fax: 86-10-80488676						
	Email: chenchen@genomics.org.cn						
	Rice genomic sequence.						
	Class: exon-trapped.						
FEATURES							
source	1..1680						
	/organism="Oryza sativa (indica cultivar-group)"						
	/mol_type="genomic DNA"						
	/db_xref="taxon:39946"						
	/clone_lib="Oryza sativa Express Library"						
	/note="Oryza sativa exon trapped genomic sequences"						
ORIGIN							
Query Match	3.2%; Score 79.6; DB 9; Length 1680;						
Best Local Similarity	41.9%; Pred. No. 0.00042;						
Matches	612; Conservative 0; Mismatches 844; Indels 6; Gaps 2;						
QY	270	CAGCGAGAGAACCGCGCCCAAGCCACACGAGCGCGAGCTGCAGGTGCGGCGGACAA	329	Db	129	CATCAACGTGGACCGGGACCATGTCCCGTTCAGACGCGGAGTTCGTACCGCA	188
QY	330	CCCCCGCAGCG	389	Db	189	CGACGCGCGCGAGACCGACCTCGACCTCGGCCACTACGAGCGTTTTCATCGAAACGCGCAT	248

```
Db 1329 CATGCGCTGGCGCGCGAAGCTCCGACGTCGCTCCGCGCACGCTGGCGCACGAGATCTA 1388
Qy 1527 GACCGACATCGTGCCCTCGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAGACCGGAGAT 1586
Db 1389 CGGCAACATCGTACCGAGCGGCCACCGCCACCGGTACGAGGCGAAGCTGAATCACTCTCGA 1448
Qy 1587 CCTGGCGAGCCCGTGCACGCGGTGTACTACGACCCCGAGCAAG---GACCTGGTGGCCGA 1643
Db 1449 CCAGCTCCGCAAGCGCGGCGCTCGTGATCTCCGCGCTCACGCGAGCGCAACAGCTCACCGA 1508
Qy 1644 GATCAGAGCAGGCGCGACGACGACCTACAGATCTACGAGGAGCCCTTCAAGAA 1703
Db 1509 GATGTCGAGCTGCGCGCGAGGTCATCTCGTGTTTCATCGCGGTGTCAGTTCACACCCCGA 1568
Qy 1704 CCTGAAGACCGCGCAAGTACGCC 1725
Db 1569 GTTCAGTCCACCCCGTGGAC 1590

RESULT 12
BM321430
LOCUS
DEFINITION
Mastigamoeba balamuthi 0.1153 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
sequence.
ACCESSION
BM321430
VERSION
BM321430.1 GI:18055836
EST.
SOURCE
Mastigamoeba balamuthi
Mastigamoeba balamuthi
ORGANISM
Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE
1 (bases 1 to 867)
AUTHORS
Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W.,
Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
Philippe, H.
TITLE
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE
21819461
PUBMED
11830664
COMMENT
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockefeller.edu
Insert Length: 867 Std Error: 0.00
POLYA=No.
Location/Qualifiers
1..867
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"

FEATURES
source
1..867
Query Match 3.2%; Score 79; DB 4; Length 867;
Best Local Similarity 45.7%; Pred. No. 0.00051;
Matches 314; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

Qy 58 CAGCGCAGCACTTCAAGGGCCCCAAGCGGCATCATCAAGTGTTCACACTGGCGGCAAGGAG 117
Db 131 CCGCACAAGATGCGGAGTGCTCGCGGTATCATCTCTGTCGCCAACAGTTGAAGTAC 190
Qy 118 GGCCACATCGCCCGCAACTGCGGCGCCCCCGCAGAGGGCTGCTGGAAGTGGCGCAAG 177
Db 191 GCGCTGACCCCGCTGAGGTACCTCGATCTGATGCGCGCTGATCAAGTACGACGGC 250
Qy 178 GAGGGCCACCGATGAAGGACTGACCGAGCGCCAGCGCAACTTCTTCGCGGAGGACCTG 237
Db 251 AAGGTCCGCGCCGACAGCGCACTTCCCGCGGGCTTCATGGACGTCGTCTCGATCGACAAG 310
```

```
Qy 238 GCCTTCCCCAGGCGAAGGCCCGGAGTTCCCGAGCAGCAGAAACCGCGCCAAACAGCCCC 297
Db 311 ACCGACGAGCACTTCCGCTCTCTAGACACCAAGGCGCGCTTCCAGGCGCACCGCATC 370
Qy 298 ACCAGCGCGAGCTGCGAGGT---GGCGGCGCAACCCCGAGCAGCGCGCGCGCCGAG 354
Db 371 AACTCGACCGAGGCGCAAGTTCAAGCTCGGCAAGGTCCGCGCGTGCAGCTCGGCAACAAG 430
Qy 355 CGCCAGGCGCACCCCTGAACTTCCCGCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATC 414
Db 431 GGCATCCCGTACCTGTGTGACCCAGACGCGCGCACGATCCGCTACCCCAACCCCGACATC 490
Qy 415 AAGTGGCGCGCCAGATCAAGGAGGCCCTCTGTGTGACACCGCGCGCGAGCACACCGTGTG 474
Db 491 AAGTCAACGACACGCTCAAGATCGACCTCGCTCGGCAAGATCATCGATTCGTTCAGG 550
Qy 475 GAGGAGATGACCTGCGCGGCAAGTGGAGCCCAAGATGATGCGCGGATCGCGCGCTTC 534
Db 551 TTCGAGATCGGCAACCTCGTTCATGATCACTGGCGGACGCAACCTTGGCGGTTCGCGCTC 610
Qy 535 ATCAAGTGGCGCAGTACGACCATGATCTGATCGAGATCTCGGCGCAAGGCCATCCGC 594
Db 611 ATTGTGCGCGCGAGAGACGAGGGCTCGTTCGAGATCATCCAGTCAAGGACGCCGCTC 670
Qy 595 ACCGTGTGTATCGCGCCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTG 654
Db 671 GGCACCAAGTTCGCGCGCGCTGACCAACGCTCTCGTGTGCGCAAGGCCACCAAGTCC 730
Qy 655 GGTGTCACCTTGAATTTCCCATCAGCCCCATCAGACCGCTGCGCGTGAAGCTGAAGCCC 714
Db 731 CTCGTACAGCTGCGCGCGCAAGGCATCAAGAAGTGCATCATCGAGGAGTTCAGGCG 790
Qy 715 GGCATGACGCGCCCCAGGTGAAGCAG 741
Db 791 GGCACCGCGCACAGGACCGAGCAGGAG 817

RESULT 13
CL978463
LOCUS
DEFINITION
OsIFCC031917 Oryza sativa Express Library Oryza sativa (indica
cultivar-group) genomic, genomic survey sequence.
ACCESSION
CL978463
VERSION
CL978463.1 GI:52411427
GSS.
SOURCE
Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Spermatophyta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Ehrhartoideae; Magnoliophyta; Liliopsida; Poales; Poaceae;
Oryza.
1 (bases 1 to 1725)
Me, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G.K.S., Deng, X.W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
Location/Qualifiers
1..1725
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"

FEATURES
source
1..1725
```



Db	1713	CCGCGCAGGGAGCAGGTGGACGCCCGCGAARACAGCTGGAGCGGTACGCTCTACACATCAA	1772
Qy	2163	CGCCTGGGCATCATCCAGGCCAGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCATAG	2222
Db	1773	GAACACGCTCGGCGGCAAGATGGCGGACGCCATGGAGGCGGAGAGAACAGAGGTGGA	1832
Qy	2223	CATCGAGCAGCTGATCAAGAAGAGAGAGGTGTACCTGAGCTGGGTGCCGCCCCACAAAGG	2282
Db	1833	GGAGGCGTGAAGGAGCGTACGAGTGGCTGGACCGCAACCCGAGCGCCGCAAGGAGGA	1892
Qy	2283	CATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAAGCAAGGGATCCGCAAGGTGCTGTT	2342
Db	1893	GTCAGAGGAAGCTGAGGAGAGCTGGAGGACGTGTGCAACCCCGTCATGTCGCGCGTCTA	1952
Qy	2343	CTTGGACGGCATCGATGGCGGCATCG	2368
Db	1953	CCAGAGTCCGGCGGCGCGCGCGCG	1978
RESULT 15			
CL952258			
LOCUS	CL952258	2697 bp	DNA linear GSS 21-SEP-2004
DEFINITION	OsJRU000500	Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.	
ACCESSION	CL952258		
VERSION	CL952258.1	GI:52364267	
KEYWORDS	GSS.		
SOURCE	Oryza sativa (indica cultivar-group)		
ORGANISM	Oryza sativa (indica cultivar-group)		
REFERENCE			
AUTHORS	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Zhao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.		
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis		
JOURNAL	Unpublished (2004)		
COMMENT	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.		
FEATURES			
SOURCE	1..2697		
	/organism="Oryza sativa (indica cultivar-group)"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:39946"		
	/clone_lib="Oryza sativa Express Library"		
	/note="Oryza sativa exon trapped genomic sequences"		
ORIGIN			
Query Match	3.2%	Score 78.2; DB 9; Length 2697;	
Best Local Similarity	43.4%	Pred. No. 0.00077;	
Matches	680; Conservative	0; Mismatches 863; Indels 24; Gaps 6;	
Qy	193	AAGACTGCACCGCGCGCAGGCGCACTTCTTCGCGGAGGACCTGGCTTCCCCAGGCG	252
Db	1141	ARGACTACCGGTGAGGCTGAGCAACGACACCGCGCAGCAGCATCTGTGGGTGGCGGTG	1200
Qy	253	AAGCGCCGCGAGTTCGCCAGCGAGCAGAAACCGCGGCCAACAGCCGCCACCGCCGAGCTG	312
Db	1201	CACCCCTCGGTGATGCTCCGCCGCGATTCTACGACGCATCTCTAACGCCCTCGAGGTG	1260
Qy	313	CAGGTGCGCGCGACAAACCCCGCAGCGAGGCGCGCGCGCAGCGCCAGGGCACCCCTGAAC	372
Db	1261	TTCAAGGTGAACACACCGCGCGCAGCTGGCGCTCGCGGACCCCGTCCCGTACAAAGCTG	1320



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 07:40:05 ; Search time 7856 Seconds  
(without alignments)

11933.846 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463

Sequence: 1 gtcagccaccatggccga.....gggctagcaccgtgaattc 2463

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34233544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_ges1:\*

9: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	136.4	5.5	330	9	CL293849
2	98.4	4.0	2886	9	CL967755
3	86.8	3.5	2598	3	AY103647
4	86.6	3.5	1398	9	CL961989
5	85.6	3.5	951	4	BM321451
6	85.4	3.5	2031	9	CL974989
7	84	3.4	1725	9	CL978463
8	80.8	3.3	3069	9	CL973991
9	80.6	3.3	869	7	CK159167
10	80.6	3.3	1941	9	CL971508
11	80.2	3.3	1132	4	BM320864
12	79.6	3.2	1165	4	BM320900
13	79.6	3.2	11691	9	CL962901
14	79	3.2	867	4	BM321430
15	78.6	3.2	2682	9	CL969033
16	78.2	3.2	2853	9	CL974397
17	78	3.2	1485	9	CL970981
18	78	3.2	2313	9	CL982362
19	78	3.2	2454	9	CL975440
20	77.4	3.1	1509	9	CL959255
21	77.4	3.1	2559	9	CL982027
22	76.6	3.1	1550	4	BM321022
23	76.4	3.1	2028	9	CL979437
24	76.2	3.1	1401	9	CL962721

25	76	3.1	1680	9	CL982770
26	75	3.1	2793	9	CL948321
27	75	3.0	862	4	BM321023
28	75	3.0	1734	9	CL948553
29	74.6	3.0	1410	9	CL947314
30	74.6	3.0	1989	9	CL980973
31	74.4	3.0	914	8	BZ568300
32	74	3.0	1290	9	CL972679
33	74	3.0	2072	3	CR603312
34	73.2	3.0	853	4	BM321393
35	73	3.0	2151	9	CL972100
36	72.8	3.0	586	4	BM587428
37	72.6	2.9	788	6	CB643171
38	72.4	2.9	753	9	CC675888
39	72.2	2.9	602	7	CV057146
40	72.2	2.9	640	2	BE601575
41	72.2	2.9	688	6	CB648640
42	72.2	2.9	764	6	CB651670
43	72.2	2.9	766	6	CB642928
44	72.2	2.9	809	6	CB641397
45	72.2	2.9	841	6	CB651502

## ALIGNMENTS

RESULT 1  
LOCUS

CL293849/c

DEFINITION

02S0349-08A1-C03 UniformMu MUTAIL Library zea mays genomic clone

CL293849

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Ze mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoidae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 330)

Authors

Latshaw S., Tan, B.-C., Settles, A.M. and McCarty, D.R.

Sequence tagged transposon insertions from the UniformMu maize population

Unpublished (2003)

Journal

Comment

Contact: Donald R. McCarty

Plant Molecular and Cellular Biology Program

University of Florida

PO 110690 Gainesville, FL 32611-0690, USA

Tel: 352-392-1928 x322

Email: drmc@ufl.edu

Sequence flanking probable Mu insertion site in UniformMu line:

02S0349-08, Primer set: A

Class: transposon insertion site.

Location/Qualifiers

1. .330

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="W22 (ACR, bz1-m9)"

/cultivar="UniformMu"

/db\_xref="taxon:4577"

/clone="02S0349-08A1-C03"

/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon

insertions in Mu inactive lines were extracted from the

UniformMu maize population by the thermo asymmetric

interlaced PCR (TAIL) protocol using primers specific for

the Mu terminal inverted repeat and a set of 16 arbitrary

primers. Amplicons were size enriched using Sepharose 400

spin columns and cloned into the TOPO PCR4 vector."

ORIGIN

Query Match 5.5%; Score 136.4; DB 9; Length 330;

Best Local Similarity 66.4%; Pred. No. 2.2e-14;



Matches	211;	Conservative	0;	Mismatches	106;	Indels	1;	Gaps	1;
Qy	421	GGGGCCGAGATCAAGGAGGGCCCTCTGGACACCGGCGCCGACACACCGCTGCTGGAGGAG	480						
Db	329	GGGGGGCAGCTGGAGAGCTCTATTAGATACAGGAGCAGATGATCAGTATTAGAGAA	270						
Qy	481	ATGAGCTGCCCCGGCAAGTGGAGCCCAAGATGATCGGCGCATCGCGCGCTTCATCAAG	540						
Db	269	ATGACTTTTGA CAGGAAGATGMAAACCAAAAATGATAGGGGAAATTGAGGGTTTTATCAA	210						
Qy	541	GTGGCCGATACGACAGATCCTGTATCGAGATCTGCGGCGAAGAGGCCATCGGCACCGTG	600						
Db	209	GTAAGACAGTATGATCAGTACAGTACCATAGAAATCTGTGGGCATAAAGCTATAGSTACGGTA	150						
Qy	601	CTGATCGGCCCCACCCCGTGAAATCATCATCGGCGCAACATGCTGACCCAGCTGGGGCTGC	660						
Db	149	TTAGTAGGACCTACACCTGTCAATTAATTGGAGAAATCTGTTGATCTAGATTGG-TGC	91						
Qy	661	ACCTGAACTTCCCCCATCAGCCCCCATCGAGACCGTGCCCGGTGAAGCTGAAAGCCCGCATG	720						
Db	90	ACCTTAAATTTTCCCATTTAGTCTTATTGAAATGTTACCAAGTAAATTAAGCCAGGAATG	31						
Qy	721	GACGGCCCCAAGGTGAAG	738						
Db	30	GATGGGCCAAAAGTAAGG	13						

RESULT 2	CL967755	CL967755	2886 bp	DNA	linear	GSS 21-SEP-2004
LOCUS	CL967755	OsiFC0015718	Oryza sativa	Express Library	Oryza sativa	(indica)
DEFINITION	CL967755	(culturivar-group)	genomic,	genomic survey	sequence.	
ACCESSION	CL967755.1	GI:52390149				
VERSION	GSS.					
KEYWORDS	Oryza sativa	(indica)	culturivar-group)			
SOURCE	Oryza sativa	(indica)	culturivar-group)			
ORGANISM	Eukaryota:	Viridiplantae:	Streptophyta:	Embryophyta:	Tracheophyta;	Tracheophyta;
	Spermatophyta;	Magnoliophyta:	Liliopsida;	Poales;	Poaceae;	
	Eriarthroideae;	Oryzaceae;	Oryza.			
REFERENCE	1.	(bases 1 to 2886)				
AUTHORS	Ma, L.,	Wang, J.,	Chen, C.,	Liu, X.,	Su, N.,	Li, L.,
	Jiao, Y.,	Sun, N.,	Zhang, X.,	Bao, J.,	Sun, D.,	Zhao, H.,
	Wong, G.K.S.,	Deng, X.W.	and	Wang, J.		
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis					
JOURNAL	Unpublished (2004)					
COMMENT	Contact: Chen Chen					
	Department of Bioinformatic					
	Beijing Institute of Genomics					
	Chinese Academy of Sciences					
	Beijing 101300, China					
	Tel: 86-10-80481559					
	Fax: 86-10-80488676					
	Email: chenchen@genomics.org.cn					
	Rice genomic sequence.					
	Class: exon-trapped					

```

FEATURES
source
1 2896
Location/Qualifiers
/organism="Oryza sativa" (indica cultivar-group)
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences"
ORIGIN

```

	Query Match	4.0%	Score 98.4;	DB 9;	Length 2886;
	Best Local Similarity	43.2%;	Pred. No. 1.7e-07;		
	Matches 709;	Conservative 0;	Mismatches 906;	Indels 27;	Gaps 4;
Qy	189	GATGAAGGACTGCAACCGAGGCCACAGGCCCACTTTCTCCGAGAGCACTGGCTTTTCCCCCA	248		
Db	219	GGTCATGGAGGCGCGCCCATCATGGCCATCGTCTCGCAACGGCGCGCGAGCGCGCC	278		

Qy		249	GGGCAAGGCCCGCGAGTTTCCCGACGAGCAGAAACCGGCGCAACAGCGCCACACGACCGCGGA	308
Db		279	CGACTGCAGGCACTTCGTTCGGCATCGTCGTCTCTCTCGTCATCAACTCCAACCATCAGTTTT	338
Qy		309	GCTGCAGGTGCGC---GGCGACAACCCCAGCAGAGGCGCGCGCAGCGCCAGGSCAC	365
Db		339	CATCGAGAGAAACATGTCGGGAACCGCGCGGGCGCTCATGCGCGCTTCGCGCCCAA	398
Qy		366	CCTGAATTCCCCCAGATCACCTTGTCGACGCCCCCTTGTTGAGCATCAAGTGTGGCGG	425
Db		399	GACCAGGTGCTCAGGACGGGAAATGGCAGGAGCAGCGCTCCATCTCTGTTCCCGG	458
Qy		426	CCAGATCAAGAGGCCCTGCTGACACACCGGCGCGAAGCAACACGTGTGGAGAGATGAG	485
Db		459	CGACATCATCAGCATCAAGTCGCGCAGCATCATCCCGCGCAGACGCGGGCTGTCGAGGG	518
Qy		486	CCTGCCCGCAAGTGGNAAGCCCAAGATGATCGGCGGCATCGGCGCTTCATCAAGGTGCG	545
Db		519	CGACCGCTCAAGGTGACACAGCGGGCGCTCACCGCGAGTCGATGCCGTCAACANGCA	578
Qy		546	CCAGTACGACCATCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCGTGTGTCAT	605
Db		579	CGCGGGCAGGGCGTCTTCTCCGGTCCACCGTCAGCAGGSGCAGATCGAGGCCGTGCT	638
Qy		606	CGGCCCCACCCCCTGTAACATCATCGGCGCGAACATGCTGACCCAGCTGGGTGACCCCT	665
Db		639	CATCGCACCGCGTGCACACTTCTCGGCAAGGGCGGCACCTGCTGGACAGCACCAA	698
Qy		666	GAATTTCCCATCAGCCCCATCAGACCGTGGCCGTGAGCTGAAGTGAAGCCCGCATGACCG	725
Db		699	CAACATCGGCCCATCTTCAGCTCTGTCTCAGCGGCATCGGCAATTTCTGCATCATCTCCAT	758
Qy		726	CCCCAAGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGACGCCCATCTG	785
Db		759	CGGGTCCGATGATAATCGAGATCATGTTATGTACCCGATCCAGCACCGCGGTACCG	818
Qy		786	CGAGGAGATCGAAGAGGCGCAAGATCACCAAGATCGGCCCGCGAAGCCCTTACAACAC	845
Db		819	CGACGGCATCGACAACTGTCTCTCTCTCATCGGCGGCATCCCCATCGCCATGCCAC	878
Qy		846	CCCCGTTCGCCATCAAGAAGAAGNACAGCACCAAGTGGCGCAAGCTGGTGACATTCG	905
Db		879	CGTGTCTCGTCAACATGGCCATCGGCTCCACCGGCTGTCGAGCAGGGCGCCATCAC	938
Qy		906	CGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCGC	965
Db		939	CAAGCGGATGACCGCCATCGAGGAGATGGCGCGCATGGAGCTGCTGTGCAGCGCAAGAC	998
Qy		966	CGGCTGAAGAAAGAGAGCGTAGCCGTGTGGACGTGGGCGACGCCTATTTCAGCGT	1025
Db		999	CGGACCGCTGACGCTCAACAAGCTCAACCGTCGACCAAGAAGCTGATCGAGGTGTACGGGG	1058
Qy		1026	GCCCTGGACGAGGACTTCGCAAGTACACCGGCTTCACCATCCCGCAGCATCACACAGA	1085
Db		1059	GGGGCTGCAAGGACTTCGGTGTCTCTGTACGGCGGAGGGCGTTCGCGCTCGAAGACCA	1118
Qy		1086	GACCCCGGCATCCGCTTACAGTACAACGTGTGCCCCAGGGCTGGAAGGGCAGCCCCAG	1145
Db		1119	GGAGCCATTGACATGATCGTGGGCACTGCTCGCGACCCCAAGAGAGGCCCGCGCGG	1178
Qy		1146	CATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGCCCGCAACCCCGAGAT	1205
Db		1179	CATCAAGAGGTCCACTTCTCCGTTTC---AACCCTGTCGAGAAGCGCACGGCATCAC	1235
Qy		1206	CGTGATCTACAGGCCCTCTGTACGTGGGCAGGACCTTGAGATCGGCGCAGCACCGCGC	1265
Db		1236	CTACATCGACGGCAATGGCGAATGGCACAGGATCAGCAAGGGGCGCGCGAGCAGATCAT	1295
Qy		1266	CAAGATCGAGGAGTGGCAAGCACCTGTGCGCTGGGGCTTCACACACCCCGCACAGAA	1325
Db		1296	CGAGCTGTGAAGATGAGCAAGACGCCGGAAGAAGGTGCACACGCTGATGCACAGTA	1355
Qy		1326	GCACCAAGAGGACCCCTCTCTGTGTGATGGGTCTAGAGCTGCACCCGACCAAGTGGAC	1385





Qy 481 ATGAGCTGCCCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAG 540  
 |||||  
 Db 631 AGGATCTCGTCCGTCAAGCCATGTCGCGCTCTCGCGAGGCCACCGCGCGAGCTCGCC 690  
 |||||  
 Qy 541 GTGCGCAGTACGACGACATCTGATCGA---GATCTGCGGCAGAGAGGCGCATCGGCACC 597  
 |||||  
 Db 691 GAGTCTGCTACGCGCTGGGCAAGGATCCAGGATCGCGCAGATTCTCAACGCCAGC 750  
 |||||  
 Qy 598 GTGCTGATCGGCCCGCCACCGCTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGC 657  
 |||||  
 Db 751 GTGGGATTCGGTGTCTCTGCTTCAGAGAGACATCTCAACCTGCTG----- 798  
 |||||  
 Qy 658 TGCACCTGAACTTCCCATCAGCCCCATCGAGACCCTGCGCGTGAAGCTGAAGCCCGGC 717  
 |||||  
 Db 799 TACATCTGCGAGTGAATGGCTCTCCCGAGTGGCCAACTACTGGAAGCAGGTTATCAAG 858  
 |||||  
 Qy 718 ATGACGCGCCCAAGTGAAGCATGTGCGCCCTGACCGAGGAGATCAAGGCCCTGACC 777  
 |||||  
 Db 859 ATCAACGACTACCAAGAGAGCCGTTCTGTGAACCGCGTGTGCTCCATGTTCAACACC 918  
 |||||  
 Qy 778 GCCATCTGCGAGGAGATGGAGAGAGGCAAGATCAACAGATCGCGCCCGAGAACCC 837  
 |||||  
 Db 919 GTGCCCGGCAAGAGATCGCGTCTGCGCTTCGCTTCAGAGAGACACCGGCGACACC 978  
 |||||  
 Qy 838 TACAACACCCCGTGTTCGCCATCAAGAAGAGACACACCAAGTGGCGCAAGCTGGTG 897  
 |||||  
 Db 979 AGGAGAGCGCCGCATCGACGCTGCAAGGGGCTGATCGSCGCAAGGGCAAGGTGAGC 1038  
 |||||  
 Qy 898 GACTTCGCGAGCTGAACAGCGCACCGAGACTTCTGGGAGGTGAGCTGGGATCCCC 957  
 |||||  
 Db 1039 ATCTACGACCCCGAGGTGACGAGGACAGGTCTCAGCGCGGACCTCGCCATGAGCAAGTTC 1098  
 |||||  
 Qy 958 CACCCCGCGCGCTGAAGAAGAAGAGCTGACCGTGTGAGCGTGGCGAGCGCTAC 1017  
 |||||  
 Db 1099 GATTGGGATCACCGGTTCATCTCCAGCGATGAGCCCGACGCGCATCAAGAGGTGTCC 1158  
 |||||  
 Qy 1018 TTCAAGCTGCGCCCTGAGCAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATC 1077  
 |||||  
 Db 1159 GTGCGCTGGGAGCGCTAGGAGCGCGCAGGACGCCACCGCGGTCTGATCTCTCAACGAG 1218  
 |||||  
 Qy 1078 AACACAGAGACCCCGCGATCGCTAC 1104  
 |||||  
 Db 1219 TGGGACGAGTTTCAGGAGCTCGACTAC 1245  
 |||||

## RESULT 5

BM321451 951 bp mRNA linear EST 03-JAN-2002  
 LOCUS rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library  
 DEFINITION Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.

## ACCESSION

VERSION BM321451.1 GI:18055857

## KEYWORDS

SOURCE EST.

## ORGANISM

Mastigamoeba balamuthi

## REFERENCE

1 (bases 1 to 951)  
 Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W., Gordon,P., Durfee,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.

The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba

## JOURNAL

MEDLINE 21819461

## PUBMED

## COMMENT

Contact: Muller Miklos

Laboratory of Biochemical Parasitology

The Rockefeller University

1230 York Avenue, New York, NY 10021, USA

Email: mmuller@rockefeller.edu

Insert Length: 951 Std Error: 0.00

## POLYA=No.

## FEATURES

## source

1. .951

/organism="Mastigamoeba balamuthi"

/mol\_type="mRNA"

/strain="ATCC 30984"

/db\_xref="taxon:108607"

/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"

/note="syn: Phreatamoeba balamuthi"

## ORIGIN

Query Match 3.5%; Score 85.6; DB 4; Length 951;  
 Best Local Similarity 46.0%; Pred. No. 3.1e-05;  
 Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;  
 Qy 1728 CGCCAAGATCGGACCGCCACACCAAGAGCTGGAAGAGTGAAGCGGCGGTGCAGAA 1787  
 Db 219 CGCCAAGATCGGAGTGTGCTGCTGCAACATCTTCTGACGAGGACCAAGCGGCGGCGC 278  
 Qy 1788 GATCGCCATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCA 1847  
 Db 279 CATCGCGCAGCGCGGCTCTCGGTCTTCGCTGGAAGGGCGAGAACCTCCAGGAGTACTG 338  
 Qy 1848 GAAGGAGACCTGGAGACCTGGTGGAGCAGTACTGCGAGGCCACCTGGATGCCGAGTG 1907  
 Db 339 GGAGTGCACCTGGAAGGCCCTGTGCTTCGGGCCCTACCAAGGGCCCTCAGATCATCGTGA 398  
 Qy 1908 GGAGTTCGTGAACACACCCCGCTGGTGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1967  
 Db 399 CGACGGCGGTGACGCGACTCTGATGATTCACAGGGGTTCGCGCCGAGGACAACCCCAA 458  
 Qy 1968 CATCGGCGCGCAGACCTTCTACGTGGACGGCGCGCCCAAGCGAGACCAAGATCGGCAA 2027  
 Db 459 GCTGCTGAGAGACGACGAGGGCTCGAGGAGTTCGCTCAACACCGTCTCAAGCA 518  
 Qy 2028 GGCCTGGTACGTACCGACCGGGCGCGGAGAGATCGTGAGCTTGACCGAGACCAACAA 2087  
 Db 519 GGTCCAGAGAGGACGCGCGGCTTCTGSCACAAAGATCTCCCGAGATCCGCGGTGTGAG 578  
 Qy 2088 CCAGAGACCGAGCTGACGAGCCATCCAGCTGGCGCTGCGCGCAGCAGCGCGCAGGAGTGA 2147  
 Db 579 CGAGGAGACGACGACTGGCGTGTGAGGCTGTACAGCTGACCGCGCAGCGCAAGCTGCT 638  
 Qy 2148 CATCGTGACCGACGACGCTACGCTGGCGCATCATCCAGCGCCAGCGCCAGCAAGAGCGA 2207  
 Db 639 GTTCCCGCGCTCAACGTCAACGAC---TCTNTCACCAAGAGCAAGTTCGACACATCA 695  
 Qy 2208 GAGCGAGCTGGTGAACAGATCATCGAGCAGCTGTATCAAGAGAGAGGTGTACCTGAG 2267  
 Db 696 CGGCTGCGCGCACTCGCTCATCGACGCGCATCAAGCGCGCACCGACGTGATGCTCGCGCG 755  
 Qy 2268 CTGGGTGCGCGCCCAAGGGCATCGCGGCAAGCGAGCAGATCGACAAGCTGGTGTGACAA 2327  
 Db 756 CAAGGTGCGCGCTGCTGCGGGGTACGCGAGCTGGGCAAGGGCTGCGCGAGTCTGCTGCG 815  
 Qy 2328 GGGCATCCGCAAGGTGCTGTCTCTGACCGCATCGATGGGGGATCTGTGATCTTACCAAGTA 2387  
 Db 816 CGGCCAGGGCTGCGCGCTCATCTGTGAGGAGATCGACCCCATCTGCGCGCTGCAGGCGTC 875  
 Qy 2388 CATGAGCAGCTGTACGTGGGCGAGCGCGCCCTTAGGATCGA 2429  
 Db 876 GATGGCGCGCTTCGAGGTCAACACGCTCGAGGGGGGCTCGA 917

## RESULT 6

CL974989

LOCUS

DEFINITION

CL974989

ACCESSION

VERSION

KEYWORDS

SOURCE

CL974989 2031 bp DNA linear GSS 21-SEP-2004

OsIFCC042895 Oryza sativa Express Library Oryza sativa (indica

cultivar-group) genomic, genomic survey sequence.

CL974989

CL974989.1 GI:52404497

GSS.

Oryza sativa (indica cultivar-group)

ORGANISM *Oryza sativa* (indica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
 1 (bases 1 to 2031)  
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,  
 Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,  
 Wang, G. K. S., Deng, X. W., and Wang, J.  
 An analysis of transcriptional regulation of the rice genome and  
 its comparison to Arabidopsis  
 Unpublished (2004)

JOURNAL  
 COMMENT  
 Department of Bioinformatic  
 Beijing Institute of Genomics  
 Chinese Academy of Sciences, Beijing 101300, China  
 Tel: 86-10-80481559  
 Fax: 86-10-80488676  
 Email: chenchen@genomics.org.cn  
 Rice genomic sequence.  
 Class: exon-trapped

FEATURES  
 source  
 1..2031  
 /organism="Oryza sativa (indica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:39946"  
 /clone\_lib="Oryza sativa Express Library"  
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN  
 Query Match 3.5%; Score 85.4; DB 9; Length 2031;  
 Best Local Similarity 42.6%; Pred. No. 3.6e-05;  
 Matches 787; Conservative 0; Mismatches 1026; Indels 36; Gaps 5;

576 CGGCAAGAGGCCATCGGCACCGTGCTGATCGGCCCCACCCCGTGAACATCATCGGCG 635  
 |||||  
 153 CGGCACGAGCTACTCGTGGTGGAGTCTACCGGAACGGCCATGTCGAGATCATGCCAA 212  
 |||||  
 636 CAACATGCTGACCCAGCTGGGTGACCCCTGAACCTTCCCATCAGCCCATCGAGACCGT 695  
 |||||  
 213 CACACAGGGGAACCGGATACGCCCTCGTGGTGGCTTTCACCGACGGCGGCGGGCT 272  
 |||||  
 696 GCCCGTGAAGCTGAAGCCCGGATGAGCGGCCCAAGGTGAAGCAGTGGCCCTGACCGA 755  
 |||||  
 273 CATCGTGGCGGCCAAGAACACGCGCGGCCCAACCCGGAGCGGACCATCTACAGCG 332  
 |||||  
 756 GGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGAGATGGAAGAGAGGGCAAGATCAC 815  
 |||||  
 333 CAAAGCGGCTCATCGGCCGCGCAGTTCTCGGACGCCGAGGTGCGAGCGGACATGAAGTCT 392  
 |||||  
 816 CAAGATCGGCCCGAGNACCCCTACACACCCCGTGTTCGCCATCAGNAGNAGGACAG 875  
 |||||  
 393 GCCGTTCGCGCTGTCGACCGGAAACGGCAAGCCGACGTCGCGCTCGAGGTGAAGGACG 452  
 |||||  
 876 CACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAGCGCACCCAGGACTTCTG 935  
 |||||  
 453 CGACGTGCG-----CGTGTTCAGCCCGGAGGAGGTGAGCGCATGTGCTCAC 500  
 |||||  
 936 GAGGTGCGAGTGGGATCCCCCAACCCGCGGCTGAAGAAGAGAAGAGCGTGAACCGT 995  
 |||||  
 501 GCGGTGAAGGAGAGCGGCGGAGGCCCTACCTCGCGGAGAAAGGTTCACCCGCGCGCTCGT 560  
 |||||  
 996 GCTGGAGCTGGCGAGCGCTACTTCAGCGTGCCTCGGACGAGGACTTCGCGAAGTACAC 1055  
 |||||  
 561 CGTCCCGGCTACTTTCAGAGAGCGGAGCGGCGGCGCCACCAAGGACCGCGGGTCAATCGC 620  
 |||||  
 1056 CGCCTTTCACATCCCGCAGCATCAACAAGAGACCCCGGCGATCCGCTACCAAGTACAAGT 1115  
 |||||  
 621 CGGGCTCACCGTCGACCGCATCATCAACGAGCCCAACCGCGCGCCATCGCCTACGGCAT 680  
 |||||  
 1116 GCTGCCCCAGGGCTGGAAGGGGAGGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCT 1175  
 |||||  
 681 CGACAAGAGGGCGCGGAAGAAGAGTCTCTGCTCTTCGACCTCGCGCGGCGGACGCTTCCA 740  
 |||||

1176 GGAGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATCTACAGGCCCCCCTGTAGCTG-- 1233  
 |||||  
 741 CGTCAGCATCTCGCATCGACAACGGGTGTTTCGAGGTCTCTTGCCACAACAGCGGACAC 800  
 |||||  
 1234 ----GGCAGCGACCTGGAGATCGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGA 1289  
 |||||  
 801 CCACCTCGGCGGAGGACTTCGACCAACGCTCATGACCACTTCGTTCGTAAGGTATCCG 860  
 |||||  
 1290 CTTGCTGGCTGGGGCTTACACACCCCGACAAGAGCACCAAGAGAGAGCCCCCTTCT 1349  
 |||||  
 861 CCGAAGCACGCGCGCGACATACCGGCGACGCGCGCGCTGGGCAAGCTCCGCGGGA 920  
 |||||  
 1350 GTGGATGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCGCATTCGAGCT---GCC 1406  
 |||||  
 921 GTGCGAGCGCCCAAGCGCGCTTCAGCAACGACGACGAGTGGCGTTCGAGTTCGATC 980  
 |||||  
 1407 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAATCGGCG 1466  
 |||||  
 981 CTTGTTGACGGCTCGACTTGTTCGAGCGCTCTCCCGCGCGGTTCGAGGAGTCA 1040  
 |||||  
 1467 CAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTGTGCGGGGCGCCAA 1526  
 |||||  
 1041 CAGCGACCTCTTCAAGAAGACGATGGTCCCGTGAGGAAGCCATGGCGGACGCGCGCT 1100  
 |||||  
 1527 GGCCCTGACGACATCGTGCCCTGACCGAGAGGCGCGAGCTGGAGCTGGCGCGAGAACCG 1586  
 |||||  
 1101 GAGCAAGGCGACATCGACGAGATCGTCTCTGTCGGCGGACGACCAAGATCCCAAGT 1160  
 |||||  
 1587 CGAGATCTGCGCAGCCCGTGCAACGGCTGTACTACGACCCCGACCAAGGACCTGGTGGC 1646  
 |||||  
 1161 GCAGCAGCTGCTCAAGGACTACTTTCGGTGGCAAGGAGCCCAACCGCGCGCTCAACCCGA 1220  
 |||||  
 1647 CGAGATCAGAAGCAGGGCCACGACCATGACCTACCATCTACAGGAGCCCTTCAA 1706  
 |||||  
 1221 CGAGCCGTCGCTGACGCGCGCGCTGACGCGCCAGCATCATCAGCGGCCAGCTCGAGGA 1280  
 |||||  
 1707 GAACTGAAGACCGGCAAGTACGCAAGTGCACCGCACCGCCCAACCAACGACGTAAGCA 1766  
 |||||  
 1281 GAACAGGAGCATGATCTCTTCGAGCTCGCGCGCTCACCTTCGCGCTGGAGACGGC 1340  
 |||||  
 1767 GCTGACCGAGCCGCTGCAAGAAGATCGGCATCGGAGAGCATCTGTGATCTGGGGCAAGACCC 1826  
 |||||  
 1341 CGCGGCGTGTATGCAAGCTGATCCCGCAACACCGTGTGTCGACGAAAGACGCA 1400  
 |||||  
 1827 CAAGTTCGCTGCCATTCAGAGAGGACCTGGAGACCTGGTGGACCGCTACTTGCA 1886  
 |||||  
 1401 GGTGTTCAACGCTACAGGACCAAGACAAACGCTGACCATTCAGGTGTTTCGAGGGCGA 1460  
 |||||  
 1887 GGCCACCTGGATCCCGAGTGGGAGTTCTGTAACACACCCCGCTGTGTGAAGCTGTGTA 1946  
 |||||  
 1461 GCGGAGCATGACGAGGACCAACCGGCTGCTCGGAGGTTTCGNCTTCGCGGCGCATCGCGC 1520  
 |||||  
 1947 CCACTGGAGAAGAGGCCATCATCGGCGCCGAGACTTTCAGTGGAGCGGCGCGCA 2006  
 |||||  
 1521 GGCGCGAGGGGGCGCGCAGATCGAGTGAACG-----TTTCGAGGTGGAGCGCA 1571  
 |||||  
 2007 CCGGAGACCAAGATCGCAAGGCGGCTACGTGACCGACCGGGCCCGCGAGAAATGCT 2066  
 |||||  
 1572 CGGCATCTCAGCTGTGCGCGCGCAACAGGCCACCGGCAGGTTCGGAAGAGATCACCAT 1631  
 |||||  
 2067 GAGCTGACCGAGACCAACCAAGACCGAGCTGCAGGCAATCCAGCTGGCGCTGCA 2126  
 |||||  
 1632 CTCGGCGACGACCGCAAGATCAGCCAGGAGAGATCGACCGGATGGTTCGCGAGCGGA 1691  
 |||||  
 2127 GGACAGCGGCGAGGAGTGAACATCGTGACCGACAGCCAGTACGCGCTTGGGCAATCATCA 2186  
 |||||  
 1692 GGAGTTCGCCGAGGAGGACCGCGGCAACAGGAGCAGGTGGACGCGCCCGCAACAGCTGGA 1751  
 |||||  
 2187 GGCCCGCCCGACAGAGCGGAGCTGGTGAACCAAGAT-----CATCGACAGCT 2240  
 |||||  
 1752 GGCGTACGCTACCAACATCAAGAACACGCTCGCGGCAAGATGGCGGACGCGCATGGAGG 1811  
 |||||  
 2241 GATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGCGCGCCCGCCCAAGGGCATCGGCGCA 2300  
 |||||

Db 1812 CGAGGAGAGCAAGGTGGAGGAGCGGTGAGGAGCGGTACAGTGGTGGACGGCAA 1871  
 Qy 2301 CGAGCAGATCACAAGCTGTGTGACCAAGGGCATCCGCAAGGTGTGTCTTCTGGACGGCAT 2360  
 Db 1872 CCGGAGCGCGCAAGGAGGAGTACGAGGAGAGTCTGAGGAGCTGGAGGACGTGTGCAA 1931  
 Qy 2361 CGATGGCGGCATCGTGTATCTACCAAGTACATGAGACGACTGTACGTGGGC 2409  
 Db 1932 CCGCGTATGTGCGCGGTCTACCAAGGTCTCGGCGGCGGCGCGGC 1980

## RESULT 7

CL978463 1725 bp DNA linear GSS 21-SEP-2004  
 LOCUS OaIFCC031917 Oryza sativa Express Library Oryza sativa (indica  
 DEFINITION cultivar-group) genomic, genomic survey sequence.

ACCESSION CL978463

VERSION CL978463.1

KEYWORDS GSS.

SOURCE Oryza sativa (indica cultivar-group)

ORGANISM Oryza sativa (indica cultivar-group)

Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 1725)

Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,

Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,

Wong, G.K.S., Deng, X.W. and Wang, J.

An analysis of transcriptional regulation of the rice genome and

its comparison to Arabidopsis

Unpublished (2004)

Contact: Chen Chen

Department of Bioinformatic

Beijing Institute of Genomics

Chinese Academy of Sciences, Beijing 101300, China

Tel: 86-10-80481559

Fax: 86-10-80488676

Email: chenchen@genomics.org.cn

Rice genomic sequence.

Class: exon-trapped.

Location/Qualifiers

1..1725

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="genomic DNA"

/db\_xref="taxon:39946"

/clone\_lib="Oryza sativa Express Library"

/note="Oryza sativa exon trapped genomic sequences"

ORIGIN

Query Match 3.4%; Score 84; DB 9; Length 1725;  
 Best Local Similarity 43.7%; Pred. No. 6.3e-05;  
 Matches 470; Conservative 0; Mismatches 600; Indels 6; Gaps 2;  
 Qy 949 GGCATCCCCCAGCCCGCCCTGAAGAAGAGAGCGTGAACCGTGTGGAGCTGGGC 1008  
 Db 76 GTCTCCGCAACCCCGGAGCCCTCCACCGCATGTACCGGAGGTGAGCCTCTCGCGAC 135  
 Qy 1009 GAGCCTACTTACGCTGCGCTGGAGCAGACTTCGCGAAGTACACCGCTTACCATC 1068  
 Db 136 CTCGCGACACGCGCATCATCGCTTCCAGCGCTCTGGGTCCACCCCTCCCGCGCAC 195  
 Qy 1069 CCAGATCAACACAGAGACCCCGGCATCCGCTACCAAGTACAGTGTGCCAGGGC 1128  
 Db 196 TTCACTTATACCGAGCTCTTCTCTCGGACACCTCCGCTCTACCGCTCCGGTAC 255  
 Qy 1129 TGAAGGCGAGCCCGAGCATTTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCGC 1188  
 Db 256 CCGCGCTGAGCGCGCGCGTCCGCGCTGGCGGCGCGCATCTCTCCACCGCTCGCC 315  
 Qy 1189 GCCGGAACCCCGAGATCGTGTATCTACAGGCCCTCTGACGTGGGCGAGACTGGAG 1248  
 Db 316 TACCTCACTCCGCGCGCTCATCCACCGGACCTCAAGTGCAGACATCTTGTCAAC 375

Qy 1249 ATCGGCCAGCACCGCCCAAGATCGAGAGCTGCGCAAGCACCTGTGCTGGGGCTTC 1308  
 Db 376 GGCACCTCGGCCAGGTCAAGATCGGCGACTCGGCTCGCGCGCTCTCGGGCTGC 435  
 Qy 1309 ACCACCCCGGACA---AGAAGCACCAGAAGAGCCCTTCTGTGTGGTGGCTACGAG 1365  
 Db 436 ACCTCGCCCGCAGACGCTCATCGGACCGCGGATTCATGGCCCCCGAGATGTACGAG 495  
 Qy 1366 CTGCACCCCGACAAAGTGGAGCCGTGACGCCATCGAGCTGCCCGAGAAGAGAGCTGAC 1425  
 Db 496 TGCTACGGCTCGGCTCGAGCTTCTCTTGGCATGTGCTGCTCGAGATGCTCACC 555  
 Qy 1426 GTGAACGACATCCAGAAGCTGGTGGCAAGCTGAATGGGCCAGCCAGATCTACCCGGC 1485  
 Db 556 AACGAGTACCCCTACAGCAGTGGCAACACCGCGCAGATCTACAAGAAGGTCAACGCC 615  
 Qy 1486 ATCAAGTGGCCAGCTGTGCAAGCTGCTGGCGGCGCCCAAGGCCCTTACCGGATCG-- 1543  
 Db 616 GGCAAGCTCCCGACGCTTCTACCTCTCACCGACCGCGCCCGCTTCTCATCGGC 675  
 Qy 1544 -TGCCCTGACCGAGGAGCGGAGCTGGAGCTGCCCGAGAACCGCGAGATCTCTCGCGAG 1602  
 Db 676 CGGTGCTTGTGACGCGGCCACCGCCCTCTGCCGAGAGCTCTGCTCGACCCCTTC 735  
 Qy 1603 CCGGTGACGCGCTGTACTACGACCCCGACCAAGGACCTGGTGGCGGAGATCCAGAAGCAG 1662  
 Db 736 CTCTCCCGCTCAGAACACGAGCATCAATACCATTTGCTCATGCCACTTGGCGCCCA 795  
 Qy 1663 GGCACGACAGTGGACCTTACAGATCTACAGAGGCCCTTCAAGAAGCTTGAAGACGGC 1722  
 Db 796 CCGCATTTGCCATTGGCGTGCAGTAATTCCAGTGAGGAGCAAGAGGAGGAGGCGCG 855  
 Qy 1723 AAGTACGCCAAGATGCGCACCGCCACACCAAGAGCTGAAGAGCTGACCGAGCGCTG 1782  
 Db 856 GCAGCAAGAGCAGCGGCTATGGCGATCCCGGAGCTGAACAGGAGCAGCAGCATC 915  
 Qy 1783 CAGAAGATCGCCATGGAGAGCATCTGTATCTGGGGCAAGACCCCAAGATTCGCGCTGCC 1842  
 Db 916 TTCTCAAGTGCAGATCGCGCGCGGCAACGTGCGGAACATCTACTTCTCCGTTGAC 975  
 Qy 1843 ATCAGAGGAGACCTGGGAGACCTGTGTGACCGACTACTTGGCAGGCCACCTGTATGCC 1902  
 Db 976 GTGGCGAACGACACGCGGATGGAGTGGCGACGAGATGGTGAAGAGCTGGACATGCC 1035  
 Qy 1903 GAGTGGGAGTTGTTGAACACCCCGCTGTGAAGCTGTGTACAGCTGGAGAGAGAG 1962  
 Db 1036 GACGGGAGCCCGAGATTTGCGCCATGATCGAGCAGGAGATCTGTAGTTGGTACCA 1095  
 Qy 1963 CCATCATCGCGCGCAGACCTTCTAGTGGAGCGCGCCGCCAACCCGAGACCAA 2018  
 Db 1096 GGGTACAGCAGCAGCATGATTCTGTAGCGGATGACGACGAGATGATGAGTCAA 1151

## RESULT 8

CL973991

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

CL973991 3069 bp DNA linear GSS 21-SEP-2004  
 OaIFCC025416 Oryza sativa Express Library Oryza sativa (indica  
 cultivar-group) genomic, genomic survey sequence.

CL973991

CL973991.1

GSS.

Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 3069)

Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,

Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,

Wong, G.K.S., Deng, X.W. and Wang, J.

An analysis of transcriptional regulation of the rice genome and

its comparison to Arabidopsis







Db 345 GCAGTTCCGCGGAGGAGATCTCGTCCATGGTGCTGATCAAGATGAGGAGATCGCGGA 404  
Qy 507 CAAGATGATCGCGGCATCGCGGCTTTCATCAAGGTGCGCCAGTACGACCATCTCTAT 566  
Db 405 GGCCTACTCGGCTCGTCCATCAAGAACGCGGTGTCAACCGTCCCGGCTACTTCAACGA 464  
Qy 567 CGAGATCTCGCGCAAGAGGCGCATCGGACACCGTGTGATCGGCCCGCCACCCCGTGAACAT 626  
Db 465 CTCGCAGAGGAGGCCACCAAGGACGCGGCGTCAATCGCGGGCTCAAGTGTATCGGAT 524  
Qy 627 CATCGCCCGCAACATGCTGACCCAGTGGGCTGCAACCTGAACTTCCCATCAGCCCCAT 686  
Db 525 CATCAACGAGCCACCGCGCGCATCGCGTACGCGCTTCGACAAGAGGCGACGAGCAG 584  
Qy 687 CGAGACCGTCCCGTGAAGTGAAGCCCGGCGATGAGACGGCCCAAGGTGAAGCATGTGCC 746  
Db 585 CGCGAGAGAACGTGCTCATCTTTGACCTCGCGCGCGCACGTTTCGATGTGTCTGCTGT 644  
Qy 747 CTTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAGATGGAGAAGGAGG 806  
Db 645 CACCATCAGAGGAGGCACTTCGAGGTCAAGGCCACCGCGCGGACACCCACTCGCGG 704  
Qy 807 CAAGATCACAGATCGGCGCCGAGAACCCCTTAC---AACACCCCGCTGTTGCGCATCAA 863  
Db 705 CGAGGACTTCGACCAACCGCATGTGTAACCACTTGTGCAAGGATTCGAAGCGCAAGACAA 764  
Qy 864 GAAGAAGACAGCACCAAGTGGCGCAAGTGTGTGACTTTCGCGAGCTGAACAAGCGCAC 923  
Db 765 GAAGGACATCAGCGGCAACCGCGCGCGTTCGCGCGGCTCAGGACGGGTGCGAGCGGCG 824  
Qy 924 CCAGGACTTCTGGGAGGTGCAGCTGGGATCCCGCACCCCGCGCGCTGMAAGAAGAA 983  
Db 825 CAAAGCGAGCTGTGTCGACGGCGAGACACCAATTGAGATCGACTCGCTGTACGAGG 884  
Qy 984 GAGCGTGACGTGTGCGAGCGGCGAGCGCTTACTTACGCGTGCCTTGGAGAGGACTT 1043  
Db 885 CATCGACTTCTACACGAGATCAACAGGCGCGGTTCGAGGAGCTCAACATGGACTGT 944  
Qy 1044 CGGAAGTACACCGCTTACCATCCCGAGCATCAACACGAGACCCCGGGATCCGCTA 1103  
Db 945 CGCAAGTGCATGGAGCGCGCTGGAGAAGTGCCTCGCGACGCGCAAGATGGACAAGAGCAG 1004  
Qy 1104 CCAGTACAACAGTGTGCTGCCCGCGGCTGGAAGGCGAGCGCCAGCATCTTCCAGAGCAGCAT 1163  
Db 1005 GTGTCAGAGCTGTGCTGTGCTGCGGCGTTCACCGCGATCCCAAGGTGCAGCAGCTGCT 1064  
Qy 1164 GACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCCGAGATCGTGTACTACAGGCGCC 1223  
Db 1065 CCAGGACTTCTTCAACGGCAAGGAGCTGTGCA-----AGAGCATCAACCCCGAGGCG 1118  
Qy 1224 CTTGTACGTGGGCGAGACTGGAGATCGGCCAGCACCGCGCAAGATCGAGGAGCTGG 1283  
Db 1119 CTTGGCGTACGCGCGCGCGCTCCAGCGCGCATCTTGAAGCGGAGGCAACGAGAAGGT 1178  
Qy 1284 CAAGCACCTGTGCTGCGTGGGCTTCAACCCCGGAGAGGACCAAGAGGAGGAGGAGGCG 1343  
Db 1179 CGAGGACTTCTTCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1238  
Qy 1344 CTTCTGTGTGATGGGCTACGAGCTGACCCCGGAGAGGAGGAGGAGGAGGAGGAGGAG 1400  
Db 1239 CGTATGACCGTGTCTCATCCCGAGGAAACACCATCCCGACCAAGAGGAGGAGGAGGAG 1298  
Qy 1401 GCTGCCCGAGAGGAGAGTGAACCGTGAACGACATCCAGAGCTGGTGGGAGAGCTGAA 1460  
Db 1299 CTTCCACCTACTCCGAGCAACAGCCCGGCGTCTGATTCAGAGTGTACGAGGCGAGCGGCG 1358  
Qy 1461 CTTGGGCCAGCAGATCTACCCCGGATCAAGGTGCGCGAGCTGTGCA-----AGCT 1511  
Db 1359 CAGGACCAAGGACAAACCTCTCGGCAAGTTCAGCTCTCCGCGATCCCAAGGAGGAGGAG 1418  
Qy 1512 GCTGCGCGGCGCAAGGCGCTGACCGACATCGTGGCCCTGACCGGAGGAGGCGAGCTGGA 1571  
Db 1419 GCGCGGCGTGCCTCAGATCAAGCTCTGCTTTCGACATCGAGCGCAACGCGCATCTGAACGT 1478

Qy 1572 GCTGCCGAGAACCGCGAGATCTTGGCGAGCGCGTGCACGCGGTGTACTACGACCCCGAG 1631  
Db 1479 GTCCCGGAGGACAAAGACGAGCGGGGCAAGAAACAAGATCAACATCACCACGCAAGGG 1538  
Qy 1632 CAAAGACCTGTGGCCGAGATCCAGAAAGAGGCGCACGACCATGAGGAGCTTACAGATCTA 1691  
Db 1539 GCGCTGAGCAAGGAGAGATCGAGAAGATGGTGAGGAGGAGAGTACAAGGCCGA 1598  
Qy 1692 CAGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGGCGCACCGCCACAC 1751  
Db 1599 GGAACGAGGACAAAGAAAGAGTGGACGCGCAAGAACGCCCTCGGAGAACTACGCGCTACAA 1658  
Qy 1752 CAAAGAGTGAAGCAGCTGACCGAGGCGGTGCAGAAGATCGCATGGAGAGCATGTGAT 1811  
Db 1659 CATGCGCAACACCATCAAGGACGAGAAGATCGCTTCAAGTTCGCGCGGAGCAAGAA 1718  
Qy 1812 CTGGGGCAAGACCCCAAGTTCCGCTCCCATCCAGAAGGAGACCTGGGAGACCTGTGTG 1871  
Db 1719 GAGGATCAGAGCGCCATCGATGGCGCATCAGCTGGCTCGACACCAACAGCTCGCGA 1778  
Qy 1872 GACCGACTACTGGCAGGCGCATCTGATCCCGAGTGGAGTTCGTGAACACCCCGCCCT 1931  
Db 1779 GCGCGCAGTTCGAGGACA---AGATGAAGAGCTGGAGGGCATCTGCACCCGATCAT 1835  
Qy 1932 GGTGAAGCTGTGTACCAAGCTGGAGAAGGCCCATCATCGCGCGGAGACCTTCTACGT 1991  
Db 1836 CGCCAAAGATGTACAGGGCGCGCGCGCATGTGCGGAGGATCGAGGAGGAGCGGCC 1895  
Qy 1992 GGAAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGCGCTA 2036  
Db 1896 GCGCGCGGAGCGCGCGAGGCGCGGAGGAGTTCGACTA 1940

## RESULT 11

BM320864  
LOCUS  
DEFINITION  
Mastigamoeba balamuthi 1132 bp mRNA linear EST 03-JAN-2002  
Mastigamoeba balamuthi lambda ZAP II Library  
sequence.

## ACCESSION

BM320864  
BM320864.1 GI:18055270

## KEYWORDS

EST.

## SOURCE

Mastigamoeba balamuthi

## ORGANISM

Mastigamoebidae; Mastigamoeba.

## REFERENCE

1 (bases 1 to 1132)

## AUTHORS

Bapteste, R., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H.

## TITLE

The analysis of 100 genes supports the grouping of three highly

## JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

## MEDLINE

21819461

## PUBMED

11830664

## COMMENT

Contact: Muller Miklos

Laboratory of Biochemical Parasitology

The Rockefeller University

1230 York Avenue, New York, NY 10021, USA

Email: mmuller@rockvax.rockefeller.edu

Insert Length: 1132 Std Error: 0.00

## FEATURES

Location/Qualifiers

1..1132

/organism="Mastigamoeba balamuthi"

/mol\_type="mRNA"

/strain="ATCC 30984"

/db\_xref="taxon:108607"

/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"

/note="syn: Phreatamoeba balamuthi"

Query Match

3.3%; Score 80.2; DB 4; Length 1132;

Best Local Similarity 45.3%; Pred. No. 0.0003;  
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

Qy 139 CGCGCCCCCGCAAGAGGGTGTGGAGGTGCGGCAAGAGGGGCCACACAGATGAAGGAC 198  
Db 79 CGCGCGAGGCGAAGACGAGTACCGCGCGCCANCTGGTATCCAGGACAGACAG 138  
Qy 199 TGCACCGAGCGCAGGCCAACTTCTTCGCGAGAGACTGGGCTTCCGCCAGAGGCC 258  
Db 139 TACAACAGCCCAAGTACCGTTCGTC-----GTCCGCTTCCACCAAGGAGACATCGTC 192  
Qy 259 CGCGAGTTCGCCAGGACAGACCGCGCCACAGCCCCACAGCGCGGAGCTGCAGGTG 318  
Db 193 TGCCAGATCGCTACGCCAAGATCGAGCGGACCAATCTCTCGCGCGCGCTACTCGCAC 252  
Qy 319 CGCGCGACAAACCCCGCAGGAGCGCGCGCGCGAGCGCGGACCGCTGAATTTCCCC 378  
Db 253 GAGCTCACCCCGTTCGGGCTCAAGCTCGGCTGACCACTAGCGCGCGCTAGCGACT 312  
Qy 379 CAGATCACCTGTGGCAGCGCCCTGTGTAGCATCAAGGTGGCGGCGCAGATCAAGGAG 438  
Db 313 GGCCTGTCTGGCGCGCGGTGTGTGAAGAAGCTCAACCTCGACTCCAACTACGAGGGT 372  
Qy 439 GCCCTGTGGACACCGCGCGCGCGAGCACCGTGTGTGGAGAGATGAGCTTCCCGGCAAG 498  
Db 373 GTCAAGAAGGTCAACCGCGGAGACTTACAACTCGAGGAGCTCGACGAGCGCGCGCGGT 432  
Qy 499 TGAAGCCCAAGATGATCGCGGGGATCGCGGCTTCAACAGGTGCGCAGTACGACAG 558  
Db 433 TTCAAGGCGCTGTCTGAGCTCGGCTGTGTGCGACCTGATGGCGCGCGGTGTTCGCC 492  
Qy 559 ATCTGTATCGAGATCTCGCGCAAGAAGGCCATCGGCACCGTGTGTATGGCGCCACCGCC 618  
Db 493 GCCCTCAAGGCGATGTGCGACGGCGGCTCAAGCTTCCCC---CACAGCGAGACCGGTTTC 549  
Qy 619 GTGAACATATCGCGCGCAACATGCTGACCGAGTGGCTGCTGACCTGTAATTTCCCATC 678  
Db 550 GTCGGCTTCAACCGCGCAAGAAGGAGTCAACCGCGGCTTCCCGCAAGTATATCTTC 609  
Qy 679 AGCCCATCGAGACCGTCCGCTGAAGCTGAAGCGCGCATGGAGCGCGCCCAAGTGAAG 738  
Db 610 GCGCGCCAGCTGCGCGGTATCATGAGCTCTCAAGAGCGAGAGCGCGCGCTTCGAC 669  
Qy 739 CAGTGGCCCTTGACCGAGGAGAAGATCAAGGCCCTTGACCGCATCTGCGAGAGATGGAG 798  
Db 670 CGCCAGTTCTCGCTACGCCAAGAGGGTGTACCGCGGACATGCTCGAAGATCTAC 729  
Qy 799 AAGGAGCGAGATCAACAGATCGGCGCGGAGAACCCCTTACACACCGCGTGTTCGCC 858  
Db 730 ACCGAGGCCCAACAGCAGATCGGCGCGGACCGGACCTTCTGTCGCCAAGCGCGCTCGAAG 789  
Qy 859 ATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAG 918  
Db 790 CCCGAGGCGCCAGCCCAAGCATCTGGGCGCAAGCGCGAGGCTGAGTACCGAGCGGCAAG 849  
Qy 919 CGCACCCAGGACTTCTGGAGGTGCGAGCTGGGCGATCCCCCAGCCCGCC 966  
Db 850 AACCCTGTCGCCAGAGAAGGTCCGCTGGGCTACCCCGAGCGCCCC 897

RESULT 12  
BM320900  
LOCUS  
DEFINITION BM320900.1 GI:18055306  
sequence.  
rockefeller 0.353 Mastigamoeba balamuthi lambda ZAP II Library  
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA

ACCESSION  
VERSION BM320900  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 1165)

AUTHORS Baptiste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Seneen, C. W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H.  
TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
MEDLINE 21819461  
PUBMED 11830664  
COMMENT Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockefeller.edu  
Insert Length: 1165 Std Error: 0.00  
POLYA=No.

FEATURES  
source 1..1165  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"

ORIGIN  
Query Match 3.2%; Score 79.6; DB 4; Length 1165;  
Best Local Similarity 45.1%; Pred. No. 0.00038;  
Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3;

Qy 90 CATCAAGTGTTCNACTGCGGCAAGAGGGGCCACATCGCGGCAACTGCGCGGCCCGCC 149  
Db 46 CGTCAAGAACAAGGCGTACTTCAAGCGCTTCCAGACCCAGTTCCGTCGCGCGCGAGGG 105  
Qy 150 CAAGAAGGGTGTGGAGGTGCGGCAAGAGGGGCCACAGATGAAGACTGCACCGAGCG 209  
Db 106 CAAGACGAGTACCGCGCGGCCANCTGGTGATCCAGGACAAGAACTAACAAGCC 165  
Qy 210 CCAGGCCAACTTCTTCGCGGAGGACCTGGGCTTCCCGCAGGCAAGGCCCGAGTTCCC 269  
Db 166 CNAGTACCGTTCGTC-----GTCCGCTTCCACCAAGGAGCATCTGTCGCGAGTGC 219  
Qy 270 CAGGAGCAGAACCGCGCCCAAGAGCCCGACAGCGCGGAGTGCAGTGCAGCGGCAAA 329  
Db 220 CTAGCCCAAGATCGACGCGGACCACTCTCGCGCGCTTCTCGACGAGCTCACCG 279  
Qy 330 CCCCAGGAGGAGCGCGCGCGGAGCGCGGAGGCGCTGAACTTCCCGCAGATCACCT 389  
Db 280 CTTGCGGCTCAAGTCTGGCTGACCAACTAGCGCGCGCTTACCGGCTGCGCTGCTGCT 339  
Qy 390 GTGCGAGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGGAGATCAAGGAGGCGCTGCTGGA 449  
Db 340 GCGCGCGGTGTGTGAAGAGCTCAACTTCTTCCAGTACGAGGCTGTCAGAAGGT 399  
Qy 450 CACGCGCGCGCAGACACCGTGTGTGAGGAGATGAGCTTCCCGGCAAGTGAAGGCCAA 509  
Db 400 CAACCGCGGAGACTTACAACTGTCGAGGAGCTCGACGAGGCGCGCGCTTCAAGGCC-- 457  
Qy 510 GATGATCGGCGCATCGCGGCTTCAAGGTGCGGCGGAGTACGACGAGATCTCTGATCGA 569  
Db 458 --TGCTCGAGCTGGGCTGTGGCGCACTCGACTGGGCGCGCGCTGTTCGCGCGCTCAA 515  
Qy 570 GATCTGCGGCAAGAAGGCCATCGGCGACCTGTGTGATCGGCGCCACCGCGGTGAACATCAT 629  
Db 516 GGGCATGTGGACGCGGCGCTCAACGTCCTCCCAAGGAGCGCGCTGCTGCGCTTCAA 575  
Qy 630 CGGCGCGCAACATGCTGACCGAGCTGGGCTGACCGCTTCCCGCAAGTTCGCCCATCGA 689  
Db 576 CGGCGCAAGAAGGAGCTCAACCGCGGCTCTCCGCAAGTATCATCTTCGCGCGGCCAGT 635  
Qy 690 GACGTCGCCGTGAAGCTGAAGCGCGCATGGAGCGGCGCCCAAGGTGAAGCAGTGGCCCT 749  
Db 636 CGCGCGGTATCAAGAGCTCTCAAGGAGGAGGAGCGCGCGCGCTTCGACCGCGCTTCTC 695

Qy	750	GACCGAGGAGAGATCAAGGCCCTGACCGC	ATCTGCGAGGAGATGAGAAAGGAGGCAA	809
Db	696	CGCTTACCCCAAGGAGGGTGTCA	CGCGGACATGCTCGAGAAGATCTACCCGAGGCCCCA	755
Qy	810	GATCAACCAAGATCGGCCCGGAGAACCCCTTACA	CACCCCGTTCGCGCATCAAGAAGAA	869
Db	756	-----CAAGCAGATCCGGGCGGACCCGACCTT	CGTCCCAAGCGGCTCGAAGCCCGA	809
Qy	870	GGACAGCACCAAGTGGCGCAAGCTGGTGGACTT	CCGCGAGCTGAAACAAGCGCACCCAGGA	929
Db	810	GGGGGCCAAGCCCAAGCACTGGGGCAAGCGGAGGCTG	ACTACCGAGCGCGCAAGAACCG	869
Qy	930	CTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCG	CGCCCTGGAAGAGAGAAAGAGCGT	989
Db	870	CGTGCGCCAAGAAGTTCGGCTGGGTACCCCCAG	CGCCCCCAAGAGCCATTAATTC	929
Qy	990	GACCGTGTGACGTGGCGCAGCGCTACTTT	CAGCGTGCCTTCGGACGAGG	1039
Db	930	CGGTGCCCTGTACACCGGCATCTCCCGCTGCGG	CGCTTGCCTGCTGCGG	979

RESULT 13					
CL962901					
LOCUS	CU962901	11691 bp	DNA	linear	GSS 21-SEP-2004
DEFINITION	OsIFCC008219	Oryza sativa	Express Library	Oryza sativa (indica	
		cultivar-group)	genomic, genomic survey	sequence.	

ACCESSION	CL962901
VERSION	CL962901.1
KEYWORDS	GI=52380530
SOURCE	GSS.
ORGANISM	Oryza sativa (indica cultivar-group)
	Oryza sativa (indica cultivar-group)
	Eukarya; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryzaceae; Oryza.

**REFERENCE**  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1. Liu, X., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Ma, L., Wang, J., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Jiao, Y., Sun, N., Zhang, X., Wang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis  
Unpublished (2004)  
Contact: Chen Chen

CONTACT: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: [chenchen@genomics.org.cn](mailto:chenchen@genomics.org.cn)  
Rice genomic sequence.  
Class: exon-trapped

```

FEATURES
  Class: exon-trapped.
    Location/Qualifiers
      1..11691
        /organism="Oryza sativa (indica cultivar-group)"
        /mol_type="genomic DNA"
        /db_xref="taxon:39946"
        /clone_lib="Oryza sativa Express Library"
        /note="Oryza sativa exon trapped genomic sequences"

```

ORIGIN	Query Match	Score	DB	Length
/novel-C1y2a sativa exon trapped genomic sequences	Best Local Similarity	3.2%	79.6	11691
	Matches	43.5%	Pred. No. 0.00044	
	Conservative	0	Mismatches 739	Indels 24; Gaps 4;

	150	2991	3050
Qy	CAAGAGGGCTGCTGGAAGTCGCGCAAGGAGGGCCACGAGATGAAGCACTGCACCGAGCG		
Db	CAAGAGGGCTTTCGCAAGCAGGGGCTCAAGGTCGAGCTGGGCTGAAGATCGCGCAAGT		

DB	2391	CAGGATGGCCCTTCACACACACAGGCGCCTTGAGAGGTGCAGCTTGGCCCTGAGGATGTTGGGCGGATG	3082
QY	210	CCAGGCCAACTTCTTTCGCGAGGACCTGGCGCTTCCCCCAGGCGAAGGCCCGCGAGTTCCC	269
DB	3051	GAAACAGCGCGCGCACGCGCTGACCTGTGGCTTACACGATGCCAAGGCGGAGGCCCGCGGA	3110

D/D
3051 GAAAGACGGGCGGCAGCAGGGGATCACCTCCATTGCCAAMGGGCGAGGCCCCGCGCA 319
QY            270 CAGCGAGCAGAACC CGCGCCAAC AGCCCCCA CCAGCGCC GCAGTTCG AAGTGCGCGG GACA CA 329

Db	3111	ACTGGCGGTGGA	CAAGCTCTATCGTCTCCATCGCGCGCGTGGCCCAA	CACCAATTCGGCGCTCAA	3117
Qy	330	CCCCCGCAGGAGGAGCGCGCGCCGAGCGCCAGGCGCACCCCTGAACTTCCCCCAGATCACCCCT	389		
Db	3171	CCCCGAAAGCGTGGCCCTGAAGCTCGACGAGCGCGGCGCCATCGTGTCGATGCCGACTG	3230		
Qy	390	GTGCGACGCGCCCTGTGTGAGCATCAAGTTGGCGGCGCAGATCAAGAGAGGCCCTGTCTGG-	448		
Db	3231	CAAGACCAACCTGCGCGCGTGTGGCGGTGGGCGACGCTGGTCGCGCGCCGATGCTGTGGC	3290		
Qy	449	--ACACCGCGCCGACACACACCGTGTCTGAGGAGAGATGAGCCTCCCGCGCAAGTGGAGGCC	506		
Db	3291	GCAACAAGCCGAGGAAGAGGCGGTGGCGGTGGCGGAGCGCATCGCGGCGCAGATGCGCCA	3350		
Qy	507	CAAGATGATCGGCGGCATCGCGGCTTCATCAAGAGTGGCGCAGTACAGCACAGATCCTGAT	566		
Db	3351	CGTGAATTCAACACATCCCTTGGGTATCTACACGAGCCCGAGATCGCTGGGTGGG	3410		
Qy	567	CGAGATCTGGCGGAAG--AAGGCAATCGGCAACGTTGCTGATCGGCGCCCAACCCCGTGA	6232		
Db	3411	CCGACACGAGCAGCAGCTCAAGGCGCAGCGCTGTCAGTACAAGGCGCGCACCTTCCCGTT	3470		
Qy	624	CATCATCGGCGGCAACATGTGACCACTGGGGTGCACCTGAACTTCCCCCATCAGCCC	683		
Db	3471	CCTGGCAACGGCGCGCGCCCGTGGCTGGGCGACACAGCCGGCATGGTCAAGTTCTCTGGC	3530		
Qy	684	CATCGAGACGTTGCCCTGTAAGCTGAAGCGCGGCATGAGCGGCGCCCAAGGTGAAGCAGTG	743		
Db	3531	CGATGCGGCCACGACAGATCTCTGGCGTGCATCTGTGGGCCGATGSCCAGCGAGCT	3590		
Qy	744	GCCCTTCAACGAGGAGAGATCAAGGCCCTTGAACCGCCATCTGCGAGGAGATGAGAAAGGA	803		
Db	3591	GATCTCGAGGCGCTGTGTGGCCATGGAGTTCAAGGCGCAGCGCCGAGGACATCGCGCGCAT	3650		
Qy	804	GGGCAAGATCACCAAGATCGGCGCCGAGAACCCCTACAAACACCCCGTGTTCGCCATCAA	863		
Db	3651	CTGGCACGCCCCATCCCTCGCTTTCCGAGGCGCAACAGGAAGCGGCGCTGCCCGGCTATGA	3710		
Qy	864	GAAGAAGACAGCACCAAGTGGGCGCA-----AGCTGGTGACTTTCGCGCAGCT	911		
Db	3711	GGCCGAGCTGGCGCCCAAGGGTTTCGACGGGACCCGGGCCAGCTGGCGCGCTCGAGGC	3770		
Qy	912	GAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCGATCCCCCAACCCCGCGGCT	971		
Db	3771	GCTCGAGCGCTCGCGGACGACTGGGCGGCTTCAAGGGAAGAGCGTCCAAACGCGTCAA	3830		
Qy	972	GAAGAAGAAGAGCGTGACCGTGTGAGAGTGGGGGACCGCTACTTACAGCTGTGCCCT	1031		
Db	3831	GAAGCTCATCCACCGCCCGAGGTGGCGCGCGGCTGTATATGTACGGGGGCGCTGGGCGG	3890		
Qy	1032	GGAGGAGACTTCGCGAAGTACACCGCTTTCACCATCCCCAGCATCAACAACGAGACCCC	1091		
Db	3891	TGGCAAGAGCTTCTGATGAGACTGCTTCTTAACGCGGTGCGATCCGCGCAAGGTGGG	3950		
Qy	1092	CGGCATCCGCTACCAAGTACAACTGTGTCGCCCGAGGGCTGGAAGGGCAGCCCCAGCATCTT	1151		
Db	3951	GCTGCAATTCACGAAATTCATGCGTAGGTGCAACCGGAACCTGGCGGCGCTGCAGGGCAC	4010		
Qy	1152	CCAGACAGCATGACCAAGATCTCTGGAGCGCTTCCGCGCCCGCAACCCCGAGATCTGAT	1211		
Db	4011	CGTGAACCCCGTGAACGTGTGGGCGAGCGCATGCGCAAGCGCTACAAAGTCACTCTGCTT	4070		
Qy	1212	CTACCGGCCCCCTGTACGTGGGCGACGACCTGGAGATCGGCCAGCACCGCGCAGAGT	1271		
Db	4071	CGACGAATTCACGTGGCGGACATCAAGAGCCCATGATCTGTGACCGCC-----TGCT	4124		
Qy	1272	CGAGGAGCTGCGCAAGCACCTGTCTGCGCTGGGGCTTCCACACCCCGCAAGAAGCACCA	1331		
Db	4125	GGCGGCGCTGTTCCAGACAACGGGTGGGCTCGTCCACCACTCGAACTTCAAGCCCGACGA	4184		
Qy	1332	GAAGAGCCCCCTTCTCTGTGAGATGGGCTACGAGCTGCAACCCGACAAGTGGACCGTGCA	1391		
Db	4185	CCTGTACCCGGGCGGCTGCAACGCGACCGCATCTCCGCGGCGCATGATCTGTCAACGA	4244		







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 07:40:05 ; Search time 7875.14 Seconds  
(without alignments)  
11933.846 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469

Sequence: 1 gtgagccaccatgcccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_hic.\*

4: gb\_est3.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_gss1.\*

9: gb\_gss2.\*

\*Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	136.4	5.5	330	9	CL293849
2	97.8	4.0	2031	9	CL974989
3	97.2	3.9	2886	9	CL967755
4	91.4	3.7	1509	9	CL959255
5	90.8	3.7	2598	3	AY103647
6	86.6	3.5	1398	9	CL961989
7	85.6	3.5	951	4	BM321451
8	85	3.4	869	7	CK159167
9	84	3.4	3069	9	CL973991
10	83.8	3.4	1941	9	CL971508
11	80.2	3.2	1132	4	BM320864
12	79.6	3.2	1165	4	BM320900
13	79	3.2	867	4	BM321430
14	78.4	3.2	11691	9	CL962901
15	78.2	3.2	1962	9	CL961326
16	78.2	3.2	2853	9	CL974397
17	78	3.2	1485	9	CL970981
18	76.6	3.1	1550	4	BM321022
19	76.6	3.1	2559	9	CL982027
20	76.2	3.1	1401	9	CL962721
21	75.6	3.1	671	6	CA093222
22	75.4	3.1	2299	3	AY106831
23	75.2	3.0	854	7	CK777127
24	75	3.0	862	4	BM321023

#### RESULT 1

CL293849/c

LOCUS

DEFINITION

02S0349-08A1-C03 UniformMu MUTAIL Library Zea mays genomic clone

02S0349-08A1-C03, genomic survey sequence.

ACCESSION

CL293849

VERSION

CL293849.1

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Zea mays

REFERENCE

1 (bases 1 to 330)

AUTHORS

Latschew, S., Tan, B.-C., Settles, A.M. and McCarty, D.R.

TITLE

Sequence tagged transposon insertions from the UniformMu maize population

JOURNAL

Unpublished (2003)

COMMENT

Contact: Donald R. McCarty

Plant Molecular and Cellular Biology Program

University of Florida

PO 110690 Gainesville, FL 32611-0690, USA

Tel: 352-392-1928 x322

Email: drmc@ufl.edu

Sequence flanking probable Mu insertion site in UniformMu line:

02S0349-08, Primer set: A

Class: transposon insertion site.

Location/Qualifiers

1. .330

source

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="W22 (ACR, bz1-m9)"

/cultivar="UniformMu"

/db\_xref="taxon:4577"

/clone="02S0349-08A1-C03"

/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon

insertions in Mu inactive lines were extracted from the

UniformMu maize population by the thermo asymmetric

interfaced PCR (TAIL) protocol using primers specific for

the Mu terminal inverted repeat and a set of 16 arbitrary

primers. Amplicons were size enriched using Sepharose 400

spin columns and cloned into the TOPO PCR4 vector."

#### ORIGIN

Query Match 5.5%; Score 136.4; DB 9; Length 330;  
Best Local Similarity 66.4%; Pred. No. 2.3e-14;

#### ALIGNMENTS

CL293849 330 bp DNA linear GSS 12-FEB-2004  
02S0349-08A1-C03 UniformMu MUTAIL Library Zea mays genomic clone  
02S0349-08A1-C03, genomic survey sequence.

ACCESSION

CL293849

VERSION

CL293849.1

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Zea mays

REFERENCE

1 (bases 1 to 330)

AUTHORS

Latschew, S., Tan, B.-C., Settles, A.M. and McCarty, D.R.

TITLE

Sequence tagged transposon insertions from the UniformMu maize population

JOURNAL

Unpublished (2003)

COMMENT

Contact: Donald R. McCarty

Plant Molecular and Cellular Biology Program

University of Florida

PO 110690 Gainesville, FL 32611-0690, USA

Tel: 352-392-1928 x322

Email: drmc@ufl.edu

Sequence flanking probable Mu insertion site in UniformMu line:

02S0349-08, Primer set: A

Class: transposon insertion site.

Location/Qualifiers

1. .330

source

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="W22 (ACR, bz1-m9)"

/cultivar="UniformMu"

/db\_xref="taxon:4577"

/clone="02S0349-08A1-C03"

/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon

insertions in Mu inactive lines were extracted from the

UniformMu maize population by the thermo asymmetric

interfaced PCR (TAIL) protocol using primers specific for

the Mu terminal inverted repeat and a set of 16 arbitrary

primers. Amplicons were size enriched using Sepharose 400

spin columns and cloned into the TOPO PCR4 vector."

Matches	211;	Conservative	0;	Mismatches	106;	Indels	1;	Gaps	1
Qy	421	GGCGCCAGATCAAGGAGGCGCTGCTGGACACCGCGCCGACGACACCGCTGGAGGAG	480						
Db	329	GGGGGCGAGCTGGAAGAAGCTCTATTAGATACAGGACGAGATGATACAGATTAGAGAA	270						
Qy	481	ATGAGCCTGCCCGGCAAGTGAAGCCCAAGATGATCGGCGCATCGGCGGCTTCATCAAG	540						
Db	269	ATGACTTTGACAGGAAGATGAAACCAACCAAAATGATAGGGGAAATTGGAGGCTTTTATCAA	210						
Qy	541	GTGCGCAGTACGACAGATCCTGATCGAGATCTGCGCAAGAGGAGCCATCGGCACCGCTG	600						
Db	209	GTAAGACAGTATGATCAGGTACCCATAGAATCTGTGGGCATAAAGCTATAGGTACGGTA	150						
Qy	601	CTGATCGGCCCCACCCCGGTAAACATCATCGGCGGCAACATCTGTGACCCAGCTGGGCTGC	660						
Db	149	TTAGTAGACCTACACCTGTCAATATTTGGAGAAATCTGTGACTCAGATTGG-TGC	91						
Qy	661	ACCTGGAATTTCCCATCAGCCCATCGAGCCGTCGCGGTGAAGTGAAGCCCGGATG	720						
Db	90	ACCTTAAATTTCCCATTAGTCTCTATTGAAACTGTACAGTAAATTTAAAGCCGGAATG	31						
Qy	721	GACGGCCCCAAGTGAAG	738						
Db	30	GATGCCCAAAAGTAAG	13						
RESULT 2	CL974989	2031 bp	DNA	linear	GSS 21-SEP-2004				
LOCUS	ORF0C042895	Oryza sativa Expressed Sequence Library	Oryza sativa	(indica					
DEFINITION	CU974989	genomic, genomic survey sequence.							
ACCESSION	CU974989.1	GI:52404497							
KEYWORDS	GSS.								
SOURCE	Oryza sativa (indica cultivar-group)								
ORGANISM	Oryza sativa (indica cultivar-group)								
REFERENCE	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,								
AUTHORS	Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,								
TITLE	Wong, G.K.S., Deng, X.W. and Wang, J.								
JOURNAL	An analysis of transcriptional regulation of the rice genome and								
COMMENT	its comparison to Arabidopsis								
FEATURES	Unpublished (2004)								
source	Contact: Chen Chen								
	Department of Bioinformatic								
	Beijing Institute of Genomics								
	Chinese Academy of Sciences, Beijing 101300, China								
	Tel: 86-10-80481559								
	Fax: 86-10-80488676								
	Email: chenchen@genomics.org.cn								
	Rice genomic sequence.								
	Class: exon-trapped.								
	Location/Qualifiers								
	1..2031								
	/organism="Oryza sativa (indica cultivar-group)"								
	/mol_type="genomic DNA"								
	/db_xref="taxon:39946"								
	/clone_lib="Oryza sativa Expressed Library"								
	/note="Oryza sativa exon trapped genomic sequences"								
ORIGIN									
Query Match	4.0%;	Score 97.8;	DB 9;	Length 2031;					
Best Local Similarity	42.6%;	Pred. No. 2.2e-07;							
Matches	787;	Conservative	0;	Mismatches 1032;	Indels	30;	Gaps	4;	
Qy	576	CGGCAAGAGGCCATCGGCACCGTGTGATCGGCGCCACCCCGTGAACATCATCGCCG	635						
Db	153	CGGCACGACGTACTCGTGTGCGTCGAGTCTACCGAAGCGGCATGTCGAGATCATCGCAA	212						

Db 1281 GAAACAGGAGAGCATGATCTCTCTCGACGTGCGCGCGTCACTCGGCTTGGAGACGGC 1340  
Qy 1773 GCTGACGAGCGCGTGCAGAGATCGCATCGAGAGCATCGTGATCTGGGCAAGACCCC 1832  
Db 1341 CGCGCGGTGATGCGCAAGCTGATCCCGCGCAACACGCTGTGCGGACGAGAGACGCA 1400  
Qy 1833 CAAGTTCCGCTGCCATCAGAGAGAGACCTGGAGACCTGGTGAGACCGACTACTGGCA 1892  
Db 1401 GGTGTTTACCACGCTACAGGAGCAAGAGCAGACCAACGCTGACCATCCAGGTGTTTCGAGGGCGA 1460  
Qy 1893 GGCCACCTGGATCCCGAGTGGAGTTCGTGAACACACCCCGCTGCTGTAAGCTGTGGTA 1952  
Db 1461 GCGGAGCATGACGAGGAGCAACCGGCTGCTCGGAGGTTCGACCTCGCGGCAATCGCGCC 1520  
Qy 1953 CCAGCTGAGAGAGAGCGCCATCATCTCGCGCGCGAGACCTTCTACGTGAGACGCGCGCCAA 2012  
Db 1521 GCGCGGAGGGGCGCGCAGATCGAGGTACG-----TTCGAGGTGGAGCGAA 1571  
Qy 2013 CCGGAGACCAAGATTCGGCAAGCGCGCTAGTGAACCGACCGGGCGCGGAGAGATCGT 2072  
Db 1572 CGGCATCTCAGCGTGTGCGCGCGCAAGAGCCACCGGAGGTTCGAGAGAGATCACCAT 1631  
Qy 2073 GAGCTGACGAGACCAACACAGAGACCGAGCTGCGAGCCATCCAGCTGGCGCTTGA 2132  
Db 1632 CTCGCGGACACCGCAAGATCAGCCAGAGAGATCGACCGGATGTGCGGAGCGGA 1691  
Qy 2133 GGACAGCGGAGCGAGGTGAACATCTGACCGACAGCGAGTACCGCTTGGGCAATCCCA 2192  
Db 1692 GGAGTTCCGCGAGAGGACCGCGGACAGGAGAGGTGACCGCGGACACGCTTGA 1751  
Qy 2193 GGCCAGCGGACAAAGAGCGAGAGCTGTGTGAACAGAT-----CATCGAGAGCT 2246  
Db 1752 GCGCTAGCTCTACAACTCAAGAACACGCTCGCGGCAAGATGCGGAGCGCATGTGAGGG 1811  
Qy 2247 GATCAAGAGAGAGAGGTGACCTGAGCTGGTGGTCCCGCCACAGAGGCAATCGCGGCA 2306  
Db 1812 CGAGGAGAGAGAGAGGTGAGGAGCGGTGAGGAGCGGTACAGTGGTGGACGCGAA 1871  
Qy 2307 CGAGCAGATCACAAGCTGTGTGAGCAAGGGCATCCGCAAGGTGTGTCTTGTGACGGCAT 2366  
Db 1872 CCGGAGCGCGGAGAGGAGGTACGAGGAGAGCTGAGGAGCTGGAGGACGTGTGCA 1931  
Qy 2367 CGATGGCGGCATCGTATCTTACAGTATACGAGCATGACGACTGTACGTGGC 2415  
Db 1932 CCGCGTATGTGCGCGGTCTACAGAGTCCGCGGCGCGCGCGCGC 1980

## RESULT 3

CL967755  
LOCUS OsiFCC015718 Oryza sativa Expressed Sequence Tag (EST) Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.  
DEFINITION OsiFCC015718 Oryza sativa Expressed Sequence Tag (EST) Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.  
ACCESSION CL967755  
VERSION CL967755.1 GI:52390149  
KEYWORDS GSS.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 2886)  
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W., and Wang, J.  
An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis  
Unpublished (2004)  
REFERENCE  
AUTHORS Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559

Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.  
FEATURES  
source 1..2886  
Location/Qualifiers  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Expressed Library"  
/note="Oryza sativa exon trapped genomic sequences"

## ORIGIN

Query Match 3.9%; Score 97.2; DB 9; Length 2886;  
Best Local Similarity 43.2%; Pred. No. 2.8e-07;  
Matches 712; Conservative 0; Mismatches 903; Indels 33; Gaps 4;  
Qy 189 GATGAAGACTGACACGAGCGCCAGGCACTTCTTCGGGAGGACCTTGGGCTTCCCGCA 248  
Db 219 GGTGATGAGCGCGCGCGCCATCATGCGCATCTCTCGCCAAACGGCGCGGCGCGCC 278  
Qy 249 GGGCAAGCGCGCGAGTTCCTCCAGCGAGCAAAACCGCGCAACAGCCCGCCAGCGCGCA 308  
Db 279 CGACTGCGAGACTTTCGTCGGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 338  
Qy 309 GTGTCAGGTGCGC---GGGAGACAAACCGCGAGCGAGCGCGCGCGCGCGCGCGCGCAC 365  
Db 339 CATCGAGGAGAACTATCCCGGAAACCGCGCGCGCTCATGCGCGCGCTTGGCGCGCAA 398  
Qy 366 CTTGAATCTCCCGAGATCACTCTGTCGAGCGCGCGCTCTGTGAGCATCAAGGTGGCGG 425  
Db 399 GACCAAGGTGCTCAGGAGCGGGAATGCGAGGAGAGAGCGCGCTCTCTCTCTCTCTCT 458  
Qy 426 CCAGATCAAGAGAGCGCTGTGGAACACCGCGCGCGCGCACACCGCTGTGAGGAGATGAG 485  
Db 459 CGACATCATCAGCATCAAGCTCGCGCATCATCTCCCGCGAGCGCGCGCTGTCTGAGG 518  
Qy 486 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCTATCAAGGTGCG 545  
Db 519 CGACCGCTCAAGTGCACAGCGCGCGCTCACCGCGAGTTCGATGCGCGCTCAACAGCA 578  
Qy 546 CCAGTACGACCATCTGATCGAGATCTGCGGCAAGAGCCATCGGACCGCTGCTGAT 605  
Db 579 CGCGGCGAGGGCGTCTTCTCCGCGTCCACCGTCAAGCAGGCGGAGATCGAGGCGCGT 638  
Qy 606 CGGCGCGCACCGCGTGAACATCATCGCGCGCAACATGCTGACCGAGCTGGGCTGCACCT 665  
Db 639 CATGCGCACCGCGGTGCGACACTTCTTCGGCAGGCGCGCGCACCTGGTGGACAGCA 698  
Qy 666 GAATTTCCCATCAGCCCCATCGAGACCGTTCGCGTGAAGCTGAAGCCCGGCAATGGACGG 725  
Db 699 CAACATCGGCGACTTCCAGCTCGTGTCTCAGCGCATCGGCAACTTCTGTCATCATCTCCAT 758  
Qy 726 CCCCAGGTGAAGAGTGGCGCTGACCGAGGAGAGATCAAGGCGCTTGACCGCGCATCTG 785  
Db 759 CGCGTCCGATGATAATCGAGATCATCTGTATGTACCCGATCCAGACCGCGCGCTACCG 818  
Qy 786 CGAGGATGAGAGAGGAGGCGCAAGATCACCAAGATCGGCGCGCGCGAGACCCCTACAAAC 845  
Db 819 CGAGCGCATGACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 878  
Qy 846 CCGCGTGTTCGCCATCAAG 905  
Db 879 CGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 938  
Qy 906 CGAGTGAACAGGCGGCGCGGAGCTTCTGGAGGTGAGCTGGGCGATCCCGCGCGCGCG 965  
Db 939 CAAGCGGATGACCGCGCATCGAGGAGATGGCGCGCATGAGCTGCTCTGTCGAGCGCAAG 998  
Qy 966 CGGCTGAAG 1025  
Db 999 CGGAGCGTGAAG 1058







Db 1476 TCAGTACCTGGGCAACCCCATCACCAACCACTGTCAGAGCGCGGAGCAGCACAAACAGG 1535

Qy 1364 GCTACGAGCTGCACCCCGGACAAAGTGGACCTGTGACGCCCATCGAGCTGCGCGGAGGAGA 1423

Db 1536 ACGTGAATCCCTGGGCTGTCTCGGCGAGGAGACCGCGAGGCGATCGACATCCTGA 1595

Qy 1424 GCTGGACCGTGAACGACATCCAGAACTGGTGGGCAAGCTGAACTGGGCGAGCCAGATCT 1483

Db 1596 AGCTCATGTCTCACCTACATCTGTGGCGTGTGCCAGGCGGTGGACCTGGCGCACTCG 1655

Qy 1484 ACCCGGCAATCAAGTGGCGGACGTGTGCAAGCTGTGCGCGCGCGGAGGCCCTGACCG 1543

Db 1656 AGGAGAACTCAAGGCGTGTGTTGAAGAAACACCGTGCACCCAGTGGGCGCAAGAGGTGTGA 1715

Qy 1544 ACATGTCGCCCTGACCGGAGGCGGAGCTGAGTGGGCGGAGAA-----CGCG 1594

Db 1716 CCATGAACCCCTCGGGCGAGCTTCCAGCGCCCGCTTTCAGCGAGAAAGGAGCTGATCAGCG 1775

Qy 1595 AGATCTCGCGGAGCGGTGCAGCGGTGTACTACGACCCCAAGGACCTGGTGGCGG 1654

Db 1776 CCATGACCGGAGGCGGTGTTACGTACGCGGAGGACGCGGCGAGCGCTGCGG 1835

Qy 1655 AGATCAGAGAGGCGCACGACAGTGGACCTTACCAGATCTTACCAGGAGCCCTTCAAGA 1714

Db 1836 TGATGAGAGCTGCGCGCGTGTGTGTGACACGCGCTCAGCAGCG--GCGACGCG 1892

Qy 1715 ACCTGAGACCGCAAGTAGCCCAAGATGCGCACCGGCCACACCAAGAGCTGAGGAGC 1774

Db 1893 AGCGGAGGCGCTCCGTGTTCTCAAGATCACAGGTTTCGAGGAGGAGCTCCGCGCGGTG 1952

Qy 1775 TGACCGAGGCGGTGCAGAGATCGCATGAGAGAGCATCGTG 1815

Db 1953 TGCCCCAGAGGTGGAGGCGCGCGCGTGGCGTGGCGGAG 1993

RESULT 6

CL961989

LOCUS 1398 bp DNA linear GSS 21-SEP-2004

DEFINITION OlfIC006991 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.

ACCESSION CL961989

VERSION CL961989.1 GI:52378720

KEYWORDS GSS.

SOURCE Oryza sativa (indica cultivar-group)

ORGANISM Oryza sativa (indica cultivar-group)

REFERENCE 1 (bases 1 to 1398)

AUTHORS Ma,L., Wang,C., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.

TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis

JOURNAL Unpublished (2004)

COMMENT Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.  
Location/Qualifiers  
1. .1398  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Exprim Library"  
/note="Oryza sativa exon trapped genomic sequences"

FEATURES

source

ORIGIN

Query Match 3.5%; Score 86.6; DB 9; Length 1398;  
Best Local Similarity 45.9%; Pred. No. 2.2e-05;  
Matches 508; Conservative 0; Mismatches 569; Indels 30; Gaps 5;

Qy 1 GTCGAGCCCAACATGGCCGAGCCATGAGCCAGGCCACAGCCCAACATCTCTGTATGTCAG 60

Db 166 GTCGAGAGGACGTCGCGGAGGCGGACATCGTCTTCTCCGTGAACACCCCAACAAAG 225

Qy 61 CGCAGCAACTTCAAGGGCCCAAGCCGATCATCAAGTGTCTTCACTGCGCAAGGAGGGC 120

Db 226 GCCCGCGGCTCGCGCCCGCAAGCGCGCG---ACCTCACCTACTGGGAGAGCGCGCG 282

Qy 121 CACATCGCGCAACTGCG 180

Db 283 CGGATGATCG 342

Qy 181 GGCACACAGATGAAGACTGACCGGAGCGGAGGCCAACTTCTTCCGCGAGGAGCTTGCGC 240

Db 343 GTCAGAGCGCGGAGGCCATCGAGAAAGATCTCGACCAACCGCGCGCGCGCGCGCGCG 402

Qy 241 TTCCCCCGGGAAGCG 300

Db 403 TTCAGATCTCTTCAACCGGAGTTCCTCGCGCGGAGGACACCGCATTCGCGGACCTGCTC 462

Qy 301 AGCGCGAGCTGCAGGTGCGCGCGGACAAACCGCGCGAGCGCGCGCGCGCGCGCGCGCG 360

Db 463 GCGCCCGA-----CGCGGTCCTCATCGCGCGCGCGGAGACCGCGCGCGCGCGCG 513

Qy 361 GGCACCTGAACTTCCCCAGATCACCTGTGTGCGAGCGCGCGCGCGCGCGCGCGCGCG 420

Db 514 GCGCTCCAGCGCTCAAGGACGTGTACGCGCGGATGCGCGCGCGCGCGCGCGCGCGCG 570

Qy 421 GCGCGCGAGATCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480

Db 571 ACCACCAACTCTGTGTCGCGCGGCTGTCCAGCTCGCGCGCGCGCGCGCGCGCGCGCG 630

Qy 481 ATGAGCTTGCCTGGCAAGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540

Db 631 AGGATCTCGTCTCAACCGCATGTCCGCGCTCTGCGAGGCGCGCGCGCGCGCGCGCGCG 690

Qy 541 GTGCGCGAGTACGACAGATCTCTGATCGA---GATCTGCGGGAAGAGGCCATCGCGCAC 597

Db 691 GAGGTGCGCTACGCGCGTGGGCAAGGACTCCAGGATCGCGCGCGCGCGCGCGCGCGCG 750

Qy 598 GTCTCTGATCG 657

Db 751 GTGGGATTCGTGGCTCTCTGCTTCCAGAAAGATCTCTCAACCTGTG-----CTG 798

Qy 658 TGCACCTTGAACCTTCCCATTCAGCCCGCATCGAGACCGTGCCTCGTGAAGCTGAAGCGCG 717

Db 799 TACATCTGCGAGTGCAATGGCTCTCCCGAGGTGGCCAACTACTTGAAGCAGGTTATCAAG 858

Qy 718 ATGAGCG 777

Db 859 ATCAACGACTACCAAGAGAGCGGTTCTGTGAACCGGCTGTGCTCTCATTTTCAACACC 918

Qy 778 GCCATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGSCCGCGGAGACCCC 837

Db 919 GTGCGCGGCAAGAAAGATCGCGGTCTGCGGCTTCGCGCTTCAAGAGGAGACACCGGCGAC 978

Qy 838 TACAACACCCCGTGTTCGCCATCAAGAAGAAGAGCAGCACCAGTGGCGCAAGCTGGTG 897

Db 979 AGGGAGACCGCGCCATCGACGTCTGCAAGGGGCTGATCGGCGCAAGGCGCAAGGTGAGC 1038

Qy 898 GACTTCGGGAGTGAACAGCGCACCGCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCC 957

Db 1039 ATCTACGACCCCGAGGTGACGGAGGACCGAGGTCCAGCGCGACCTCGCCATGAGCAAGTTC 1098

Qy 958 CACCCCGCGCGCTGAAGAAGAAGAGGCGTCAACCGTGTGACGCGTGGCGCGACCGCTAC 1017

Db 1099 GATTGGATCACCCGCTTCCATCTCCAGCGGATGAGCCCGCGCGCGCGCGCGCGCG 1158

Qy 1018 TTCAGCGTGCCTGGACGAGGACTTCCGCAAGTACACCGCGCTTCCACCATCCCCAGCATC 1077

```

Db 1159 GTCCCTTGGGACGGCTACGAGCGGCGCAGGACGCCACCGGCTGTGATCCTTCAACGAG 1218
Qy 1078 AACAAACGAGACCCCGGCGATCCGTTAC 1104
Db 1219 TGGACGAGTTTCAGGAGGCTCGACTAC 1245

RESULT 7
BM321451 951 bp mRNA linear EST 03-JAN-2002
LOCUS rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
3.3.1.1), mRNA sequence.
BM321451
BM321451.1 GI:18055857
EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
REFERENCE 1 (bases 1 to 951)
AUTHORS Eukaryote; Pelobiontida; Mastigamoebidae; Mastigamoeba.
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
21819461
11830664
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 951 Std Error: 0.00
POLYA-NO. Location/Qualifiers
FEATURES
source
1. 951
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/notes="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match 3.5%; Score 85.6; DB 4; Length 951;
Best Local Similarity 46.0%; Pred. No. 3.2e-05;
Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;

Qy 1734 CGCCAAGATCGCACCGCCGCCACACACGAGCGTGAAGCAGTGCACCGAGGCGGTGCAGAA 1793
Db 219 CGCAAGATCGCTGGTGTCTGTCTGACATCTTCTCGACGAGGACCAACGCGGCGCGC 278
Qy 1794 GATCGCATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCTGCCATCCA 1853
Db 279 CATCGCGACGCGCGCTCTCGGTCTTCGCTGGAAGGCGGAGAACCTCCAGGAGTACTG 338
Qy 1854 GAAGGACACCTGGGAGACCTGTGGACGACTCTGGCGGCCACCTGGATCCCGGAGTG 1913
Db 339 GGAGTGCACTGGAAGGCCCTGTCTTCGCGCCCTACACGAGGCCCTCAGATCATCGTCGA 398
Qy 1914 GGAGTTCTGTAACACCCCGCCCTTGGTGAAGCTGTGTACAGCTGGAGAGGAGGCCAT 1973
Db 399 CGACGGCGGTGACCGGACTCTGATGATCCACAGGGGTTTCGGCGCGAGGACACCCCAA 458
Qy 1974 CATCGGCGCGGACACCTTCTACGTGGAAGCGCGCCGCAACCGCGAGACCAAGATCGGCAA 2033
Db 459 GCTGCTGGAGAGACGACGAGGCGCTTCGAGGAGGTTCGCTGCTCAACACGCTGCTCAAGCA 518
Qy 2034 GCGCGGTCTAGTACCGACCGGGCGCGGAGAGATCGTGTGAGCTGTGACCGAGACCCAA 2093
Db 519 GGTCCAGAGGAGCAGCGCGGCTTCTGGCAAGAGATCCTCCCGAGATCCCGCGGTGTCAG 578

```

```

Qy 2094 CCAGAAGACCGAGCTGCAGGCGCATTCAGCTGGCCCTGCAGACAGCGCGCAGAGTGAA 2153
Db 579 CGAGGAGACGACGACTGGCGGTNATGAGGCTGTACAGCTGCACCGCGCAGCGAGCTGCT 638
Qy 2154 CATGCTGACGACAGCCAGTACGAGCGCTGGGCGATCATCCAGGCCCGCAGCCGACAGAGCGA 2213
Db 639 GTTCCCGCGCGTCAACGTCACAGGAC---TCTNTCACCAGAGCAAGTTTCGACCAACATCTA 695
Qy 2214 GAGCGAGCTGTGAAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTCGAG 2273
Db 696 CGGTTCGCGCACTCGCTCATCGAGCGGCATCAAGCGGCGCACCGACGTGATGCTCGGCGG 755
Qy 2274 CTGGGTGCGCGCCCAAGGCGCATTCGCGCGCAACGAGCAGATCGACAAAGCTGGTGAGCAA 2333
Db 756 CAAGGTGCGCGCTGCTCGCGGCTACGCGGACGTGGCAAGGCGTGGCGCGAGTGGCTCGC 815
Qy 2334 GGGCATCCGCAAGGTGCTGTTCTGGAGCGGCATCGATGGCGGCATCGTGATCTACCAGTA 2393
Db 816 CGGCGAGGCTGCGCGCTCATCTGTGAGGAGATCGACCCCATCTGCGCGCTGCAGGGGCTC 875
Qy 2394 CATGAGCAGCTGTACCTGAGCGGCGCGGCCCTAGGATCGA 2435
Db 876 GATGGCGGCTTCGAGGTCAACAGCTCGAGGCGGCGCTCGA 917

RESULT 8
CK159167/c 869 bp mRNA linear EST 05-DEC-2003
LOCUS FGAS040564 Triticum aestivum FGAS: TaL5 Triticum aestivum cDNA,
DEFINITION mRNA sequence.
CK159167
CK159167.1 GI:38985053
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1 (bases 1 to 869)
Allard,P., Crosby,M.L., Danyluk,J., Eudes,P., Frick,M., Gaudet,D.,
Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
Penniket,C., Roach,J.L. and Sarhan,F.
Functional Genomics of Abiotic Stress In Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_este@cs.usask.ca
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [128,636].
Plate: TaL537 row: N column: 23.
FEATURES
source
1. 869
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: TaL5"
/notes="Organ: Crown; Vector: pGEM-T; SSH (suppression
subtractive hybridization) cDNA library from genotype
PI178383 cold hardened at 2 C for 21 days and 49 days
(equal amount of cDNA pooled together before subtraction,
tester) and subtracted against genotype Norstar cold

```

hardened at 2 C for 1 day (24 H) (driver). Modified Smart cDNA (Clontech) priming and non-directional cloning"

ORIGIN	Query Match	3.4%	Score 85;	DB 7;	Length 869;
	Best Local Similarity	45.4%;	Pred. No. 4.1e-05;		
	Matches 304;	Conservative 0;	Mismatches 365;	Indels 0;	Gaps 0;
QY	580	AAGNAGGCCATCGCACGGTGTCTATCGGCGCCACCCCGTGAACATCATCGGCGGCAC	639		
DB	827	ATGCGCGCGCCCGCAAAACGAGGATCACCACACGAGCGAGAACAACCAACACCCAC	768		
QY	640	ATGTGTACCCAGCTGGGTGTCACCTGAACTTCCCCCATCAGCCCCATCGAGACCGTGC	699		
DB	767	CACACACGACGAACAGGGACACAAACGACACCAACACCAACACACAAACAAC	708		
QY	700	GTGAAGCTGAAGCCCGGCATGGACGGGCCCAAGGTGAAGCATGTGGCCCTTGACGAG	759		
DB	707	ACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	648		
QY	760	AAGNTCAAGCCCTGACGGCCATCTGCGAGGAGATGGAGAGGAGGCGACAGATCACC	819		
DB	647	AACAACAACGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	588		
QY	820	ATCGGCCCGGAGAACCCCTCAACAACACCCCGTGTTCGCCATCAAGAAGAAGCAGCAC	879		
DB	587	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	528		
QY	880	AAGTGGCGCAAGTGTGTGGACTTCCGCGAGCTGAACAAGCGCAACCCAGGACTTCT	939		
DB	527	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	468		
QY	940	GTGCAGCTGGGCATCCCCACCCCGCGGCTGTGAAGAGAGAGAGCGGTGACCGTGTG	999		
DB	467	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	408		
QY	1000	GACGTGGCGACGCGCTACTTTCAGCGTGCCTTCGACGAGGACTTCCGCAAGTACACG	1059		
DB	407	GACAAACACGACACACGACAAACAACAACAACAACAACAACAACAACAACAACA	348		
QY	1060	TTCACCATCCCCAGCATCAACAACGAGACCCCGGCATTCGCTACCAAGTACAACGTGTG	1119		
DB	347	GACCACAACGACAACGACAAACAACAACAACAACAACAACAACAACAACAACAACA	288		
QY	1120	CCCAGGGCTGGAGGGCGCCCGAGCATTTCCAGAGCAGCATGACCAAGATCCTGGAG	1179		
DB	287	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	228		
QY	1180	CCCTTTCGCGCGCGCAACCCCGAGATGCTGATCTTACCAGTACATGGACGACTGTAC	1239		
DB	227	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	168		
QY	1240	GGCAGCGAC	1248		
DB	167	AACGGCAAC	159		

RESULT	9
LOCUS	CL973991
DEFINITION	OAFCC025416 <i>Oryza sativa</i> Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION	CU973991
VERSION	CU973991.1 GI:52402507
KEYWORDS	GSS.
SOURCE	<i>Oryza sativa</i> (indica cultivar-group)
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; <i>Oryza</i> . 1 (bases 1 to 3069) Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
AUTHORS	

Wong, G.K.S., Deng, X.W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and  
its comparison to Arabidopsis  
Unpublished (2004)  
Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.  
Location/Qualifiers  
1. .3069  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Express Library"  
/notes="Oryza sativa exon trapped sequences"

```

ORIGIN
Query Match 3.4%; Score 84; DB 9; Length 3069;
Best Local Similarity 42.8%; Pred.No. 6.7e-05;
Matches 536; Conservative 0; Mismatches 710; Indels 6; Gaps 2;

QY 37 ACCAGCGCCAACTCTGATCGACGCGCAACTTCAAGGGCCCCAAGCGCATCATCAAG 96
DB 625 AGCGGCATCGCAAGACGACGCTGGCGAGTACTCTACCGAGGACACGATCCCGAAG 684
QY 97 TGCCTTCAACTGCGCGAAGGAGGGCCATCGCCCGCAACTGCGCGCGCCCCCGCAAGAAG 156
DB 685 CACTTCGCCCTGCACGGCGTGGTCCAGTCCCGCGAAGTTCGCTTCGCGGACTCCTC 744
QY 157 GGTGCTGGAAGTCCGCGAAGGAGGGCCACGATGAAGGACTGCACCGGAGCGCAGGCC 216
DB 745 CACGACGTCACTCGCGCAGGCGCAACGCGCTCTCCGGCGACGACGAGTTCGGCGCGAGGC - 803
QY 217 AACTTCTTCGCGAGGACCTGGCTTCCCCCAGGGCAAGCGCCGCGAGTTCGCCAGCGAG 276
DB 804 --CGCCGACGACGCTGCGCGCGCCGGGTGGATGGAGCGGAGAGGCTCGTCAAGCGC 861
QY 277 CAGAAACCGCGCAACAGACCCCAACAGCGCGGAGCTGCAGGTGGCGGCGCGACAACCCCCCGC 336
DB 862 CGCCTCGCGCGCGCTTCAGGGGAAGAGTACTCTCGTGTGCTCGACGAGCTCGGAGC 921
QY 337 AGCAGAGCGCGCGCGAGCGCCAGGGCAACCTGAACTTCCCCCAGAGTCAACCTGTGGCGAG 396
DB 922 AGGAGGAGTGGGCTTCTTCTCGCGCGCTCCCGCGCGCTCGCGCGCAGCTGCGTG 981
QY 397 CGCCCCCTGTGAGCATCAAGTGGGCGGCGCAGATCAAGGAGGCCCTGTGGACACCGCGC 456
DB 982 CTCGTCAAGCGCAGAGTCAAGTACGCGAGGATCGGAGCAACCGTGGCGCGCGCGCGC 1041
QY 457 GCCGACGACACCCGTGCTGGAGGAGATGAGCCTGCGCGCAAGTGGAGGCCAAGATGATC 516
DB 1042 GCGCTGTGCGACCAAGGAGCTCGGGAAGCTGCGGCGGAGCAGGCGCGCGAGCTGTT 1101
QY 517 GCGGCGCATCGCGCGCTTCATCAAGGTGCGCGAGTACGACGAGATCTGTGATCGAGATCTGC 576
DB 1102 CGCCGAGGGTGTACGGCCCAACGCGAGCCGAGCGCGAAGAAAGATGGAGCAGCTCAAGAGC 1161
QY 577 GGCAGAAGGCCATCGGCAACGCTGTATCGGCGCCACCCCGTGAACATCATCGCGCGC 636
DB 1162 CTCGTGACCTCCATGACGAAGGGGCTCAAACCTGCGCGCTCAACATCGTCTGCTCGCCGA 1221
QY 637 AACATGCTGACCCAGCTGGGCTGCACCCCTGAACTTCCCATCATGACCCCATCGAGACCGTG 696
DB 1222 CTCCTTGAGTCCAAGAGAGGATGATGGAGGCGCTGTATCCATACCTCGATGATACG 1281
QY 697 CCGGTGAAGTGAAGCCCGGATGGACGGGCCCAAGGTGAAGCATGTGGCCCTTGACCGCAG 756
DB 1282 CCACCCCGGACATCGCCGAGCAGACGCGCCGAGGCGGAGCGGAGAGGAGCAGCCTC 1341

```

Qy 757 GAGAAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGGAGAAGAGGCGCAAGATCAC 816  
|||  
Db 1342 GATGATACGAGAGCGGCGCCGACATCGACGAGCAAAACGAAGAAGACGAAGAGGAAG 1401  
|||  
Qy 817 AAGATCGGCGCGGAGAACCCCTACACACACCCCGTGTTCGCGCATCAAGAAGAGACAGC 876  
|||  
Db 1402 CAGATGGGACATCGATGGAACAAGATCTGACGCTGTGATGAGAGAGCTGCCACCGAT 1461  
|||  
Qy 877 ACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936  
|||  
Db 1462 CTGAAGCGCTGCTTCTCTACTTCGCGGGTTACCGCGAGACGCGGATCTGCGCGGG 1521  
|||  
Qy 937 GAGTGCAGCTGGGATCCCCACCGCGCGGCTGAAGAAGAGAGAGGCTGACCGTG 996  
|||  
Db 1522 AAGCTGTCCGCGCTGTGGGTCCGCGAGGGTTCTATGAGGCGGAAGAACGGGACAGCGTG 1581  
|||  
Qy 997 CTGACGTGGCGGACGCTACTTTCAGGCTGCGCTGGACGAGGACTTCCGCAAGTACAC 1056  
|||  
Db 1582 GAGGAGCAGCGCGAGGAGTGCTCAAGAGCTCATCTCCCGGTGCTGCTCCAGCTCGTC 1641  
|||  
Qy 1057 GCCTTACCATCCCGAGCATCAACAAGAGACCCCGGCTATCCGCTACAGTCAACAGTG 1116  
|||  
Db 1642 GAGACGAGCGCGCGGCGGAGGTCGCGCGCTCAGCATCCACCAAGCGGTCTCGAC 1701  
|||  
Qy 1117 CTGCCCCAGGCTGGAGGAGGAGCGCCAGCATCTTCCA---GAGCAGCATGACCAAGATC 1173  
|||  
Db 1702 TTCTCCAGGCGGAGGCGCGACACCAACTTCTCCACAGTGACAGCGCGCGCGCGGC 1761  
|||  
Qy 1174 CTGGAGCCCTTCGCGCGCGCAACCCGAGATGTGATCTACAGTACATGAGCAGCTG 1233  
|||  
Db 1762 CTCTCAACGCGCGCGCGCGCTCGCGCTCCGCAACACTACGACACCGACTCGCC 1821  
|||  
Qy 1234 TAGCTGGGACGAGCTGGAGATCGGCGAGCAGCAGCGCGCAAGATCGAGGAGC 1285  
|||  
Db 1822 GTCATGCTCGAGGCGCCAGCTCCACCTCTCTGCGACATCCCGAGC 1873  
|||

RESULT 10  
CL971508  
LOCUS  
DEFINITION OaIFC021485 Oryza sativa Express Library Oryza sativa (indica  
cultivar-group) genomic, genomic survey sequence.  
CL971508  
CL971508.1 GI:52397596  
GSS.

Oryza sativa (indica cultivar-group)  
Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 1941)

REFERENCE  
AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,  
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,  
Wong, G.K.S., Deng, X.W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and  
its comparison to Arabidopsis  
Unpublished (2004)

COMMENT  
Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.

FEATURES  
source  
1..1941  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Express Library"

/note="Oryza sativa exon trapped genomic sequences "

## ORIGIN

Query Match 3.4%; Score 83.8; DB 9; Length 1941;  
Best Local Similarity 42.9%; Pred. No. 7e-05;  
Matches 769; Conservative 0; Mismatches 992; Indels 30; Gaps 6;

Qy 267 CCCAGGAGAGAACCGCGCAACAGCCCAACAGCCGCGAGCTGCAAGTGGCGGCGA 326  
|||  
Db 165 CGCGCCAAAGAACAGGTTCGCGATGAACCCCAACACACCTCTTCGATGCGAAGCGTT 224  
|||  
Qy 327 CAACCCCGGAGCGCGCGCGCGAGCGCGCGGACCCCTGAACTTCCCCAGATCAC 386  
|||  
Db 225 GATTGGTAGAGTTCTCCGACCCGTCGATGACATGATGATGATGATGATGATGAT 284  
|||  
Qy 387 CCTGTGCGAGCGCCCTCTGTGTGATCAAGGTGGCGGCGAGATCAAGGAGGCGCTCT 446  
|||  
Db 285 GGTGTGCTGCGCCCGCGAGACCGCATGATCGTCCAGTACAAGGGGAGAGAA 344  
|||  
Qy 447 GGAACCGGCGCGAGCAACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAGGCC 506  
|||  
Db 345 GCAGTTCGCGCGGAGGAGATCTCGTCATGTCTGATCAAGATGAGGAGATCGCGGA 404  
|||  
Qy 507 CAAGATGATCGGCGCATCGCGCTTTCATCAAGGTGGCGGCGAGTACGACAGATCTGAT 566  
|||  
Db 405 GGCCTACTCGGTCTGCTCCATCAAGAACGCGGTGTCACCGTCCCGGCTACTTCAACGA 464  
|||  
Qy 567 CGAGATCTGGCGCAAGAAGCGCATCGGCACCGTCTGATCGGCGCCCGCCCGGTGAACAT 626  
|||  
Db 465 CTGCGAGAGCGGCGCACCAAGAGCGCGGCTCATCGCGGCTCAACGTGATGCGGAT 524  
|||  
Qy 627 CATCGGCGCAACATGCTGACCCAGCTGGGTGCACCTTGAATTTCCCGCATCAGCCCAT 686  
|||  
Db 525 CATCAACGAGCCACCGCGCATCGGTACCGCTTCGACAAAGAGGCGAGGAGCAG 584  
|||  
Qy 687 CGAGACGCTGCGGTGAAGCTGAAGCCCGGATGGAAGCGGCGGCGGCGGCGGCGG 746  
|||  
Db 585 CGCGAGAAGAAGCGTGTCTCATCTTTGACCTCGCGCGGCGGCGGCGGCGGCT 644  
|||  
Qy 747 CTTGACCGAGAGAAAGTCAAGCGCCTGACCGCATCTGCGAGGAGATGGAAGAAGGAGG 806  
|||  
Db 645 CACCATCGAGAGGCGCATCTTCAGGTCAAGGCGCACCGCGGCGACACCCACTCGCGGG 704  
|||  
Qy 807 CAAGATCACCAAGATCGGCGCGCGAGAACCCCTAC---AACACCCCGGTGTTCGCCATCAA 863  
|||  
Db 705 CGAGGACTTCGACAAACCGCATGTGAACACTTCTGTCAGGAGTTCAAGCGCAAGAACAA 764  
|||  
Qy 864 GAAGAAGGACGACACCAAGTGGCGGCGGAGCTTGGGAGCTTCGCGAGCTGAACAAGCGCAC 923  
|||  
Db 765 GAAGGACATCAGCGGCAACCGCGCGCTCGCGCGCTCAGGACGCGGCGGCGGCGGCG 824  
|||  
Qy 924 CCAGGACTTCTGGAGGTGAGCTGGGCGATCCCCACCCCGCGCTGAAGAAGAAGAA 983  
|||  
Db 825 CAAGCGACCGCTGTCTGAGCGGCGGAGACACCATTTAGATCGACTCTGCTGTACGAGGG 884  
|||  
Qy 984 GAGCGTGACCGCTGTGAGCGGCGGCGGCTACTTTCAGCGGTGCGGCTGGACGAGGACTT 1043  
|||  
Db 885 CATCGACTTCTACACGACGATCACAGGCGCGGTTCGAGGAGCTCAACATGGACCTGTT 944  
|||  
Qy 1044 CCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAGAGACCCCGGCGATCCGCTA 1103  
|||  
Db 945 CCGCAAGTGTGATGAGCGGCGGCGGAGAGTCTCTCCGCGAGCCCAAGATGGAACAAGAGCAG 1004  
|||  
Qy 1104 CCAGTACAAGGTGCTGCCCGAGGCGTGAAGAGGCGGCGGCGGCGGCTTCCAGAGCAGCAT 1163  
|||  
Db 1005 CGTGACGACGTGTGTCTC-----GTCGGCGGCTCCACCGCATCCCAAGGTGCA 1055  
|||  
Qy 1164 GACCAAGATCTGGAGCGCTTTCGCGCGCGCAACCCCGAGATCGTGTACTACCAAGTACAT 1223  
|||  
Db 1056 GCAGTGTCTCAGGACTTCTTCAAGCGCAAGGAGCTCTGCAAGAGCATCAACCCCGACGA 1115  
|||  
Qy 1224 GGAAGACCTGTACGTGGGCGAGGACCTGGAGATGGCGGCGGCGGCGGCGGCGGAGTA 1283  
|||

```
Db 1116 GGCCTGGCGTACGGCGCGCGCTGCAGGCGCCATCTTGAGCGCGAGGGCAACGAGAA 1175
QY 1284 GCTGGCAAGCACCTGCTGGCTGGGGCTTACACACCCCGCAAGAAGCACCAGAGGA 1343
Db 1176 GGTGC---AGGACCTGCTGCTGCTGACGCTCACCCGCTGTCCCTCGGCTCGAGACCGC 1232
QY 1344 GCCCCTTCTGCTGGATGGCTACGAGCTGCACCCCGCAAGTGACC---GTGAGGCC 1400
Db 1233 CGCGGGGCTCATGACCGTCTATCCGAGGAACACACCATCCGACCAAGAAGAGCA 1292
QY 1401 CATCGAGCTGCCCGAGAGAGAGCTGGACCTGTAACGAATCCAGAGAGCTGGTGGGCAA 1460
Db 1293 GGTCTTCTCCACCTACTCCGACAAACAGCCCGCGCTCTGATCCAGTGTACGAGGGCGA 1352
QY 1461 GCTGAATGGGCGAGCAGATCTACCCCGGCATCAAGTGGCGCAGCTGTGCA----- 1513
Db 1353 GCGCGCAGGACCAAGGACAACAACCTCTCGGCAAGTTTCGAGCTCTCCGGCATCCCAAC 1412
QY 1514 --AGCTGCTGCGGCGGCCAAGGCCCTGACCGACATCTGTCGCCCTGACCGAGGAGCGGA 1571
Db 1413 GGCCTCGCGGGCTGCCCCAGATCAACGTCTGCTTCGATCGAGCCCAACGGCATCCT 1472
QY 1572 GCTGGAGCTGGCCGAGAACCGGAGATCTTCGCGAGCCCGTGCACCGGCTGTACTACGA 1631
Db 1473 GAACGTGTCCGCGAGGACAAGACGACGCGGCGAGAAACAAGATCAACATCAACCAAGA 1532
QY 1632 CCCAGAGGACCTGTTGGTCCGAGATCCAGAGAGGAGGCGACGACGATGGACCTTACCA 1691
Db 1533 CAAGGGCGCGCTGAGCAAGGAGGAGATCGAGAAGATGGTSCAGGAGCGGAGAGTACAA 1592
QY 1692 GATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTAGCGCAAGATGCGCACCGC 1751
Db 1593 GGCAGGAGCAGAGAGACAAGAAGAGTGAGGAGCCCAAGAACGCTTCGAGAACTACGC 1652
QY 1752 CCAACCAACAGCTGAAGCAGCTGACCCGAGCGCTGCAGAAAGATCGCCATGGAGAGCAT 1811
Db 1653 CTACAATCGCAACACCATCAAGAGCAGAGAAGATCGCCTCCAAGTCTCGCCGCGACGA 1712
QY 1812 GTGATCTGGGGAAGACCCCAAGTTCCGCTGCCCTCAGAACGAGGACCTGGGAGAC 1871
Db 1713 CAAGAAGAGGATCGAGGACGCCATCGATGGCGCCATCAGCTGGCTCGACACCAACAGCT 1772
QY 1872 CTGGTGACCGACTACTGGCAGCGCCACTCGATCTCCGAGTGGGAGTTCGTGAACACCCC 1931
Db 1773 CGCGAGGCCAGAGTTTCAGAGACA---AGATGAAGGAGCTGGAGGGCATCTGCAACC 1829
QY 1932 CCCCTTGGTGAAGCTGTGTTACAGCTGGAGAGAGGCCCATCATCGGCGCCGAGACCTT 1991
Db 1830 GATCATCGCCAAAGATGTACCGGGCCCGCGCGGACATGGCGGAGGATGGACGAGGA 1889
QY 1992 CTAGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGCTA 2042
Db 1890 CGCGCGCGCGCGGCGAGCGCGCAGCGCCCCCAAGATCGAGGAGTTCGACTA 1940
```

## RESULT 11

```
BM320864 1132 bp mRNA linear EST 03-JAN-2002
LOCUS rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
sequence.
BM320864
BM320864.1 GI:18055270
EST.
Mastigamoeba balamuthi
Mastigamoeba balamuthi
ORGANISM Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 1132)
AUTHORS Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
```

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
MEDLINE 21819461  
PUBMED 11830664

COMMENT  
Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockvax.rockefeller.edu  
Insert Length: 1132 Std Error: 0.00  
POLYA=No.

## FEATURES

source  
1. .1132  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"

## ORIGIN

Query Match 3.2%; Score 80.2; DB 4; Length 1132;  
Best Local Similarity 45.3%; Pred. No. 0.0003;  
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;  
QY 139 CGCGCCCCCGCAAGAGGGCTCTCGAAGTGCAGGAGGAGGGCCACCAAGATGAAGGAC 198  
Db 79 CGCGCGAGGGCAAGACGGACTACCGCGCGCCANCTGGTGATCCAGGACAAACAAG 138  
QY 199 TGCAACGAGCGCAGAGCCAACTTCTTCGCGAGAGACTGTGGCTTCCCCCAGGGGCAAGGCC 258  
Db 139 TACAACAGAGCCCCAAGTACCGCTTCGTC-----GTCCGCTTCCACCAACAGGAGCATCGTC 192  
QY 259 CGCGAGTTCCTCCAGGAGGAGACCGCGCAACAGCCCCACAGCGCGAGCTGCAGGTG 318  
Db 193 TGCCAGATCGCTTACGCCAAGATCGAGCGGACACATCTCTCGCGCGCTACTCGCAC 252  
QY 319 CGCGCGCAACAACCCCGCAGCGAGCGCGCGCGAGCGCGCAGGGCACCTTGAATTTCCCC 378  
Db 253 GAGCTCACCGCTTCGCGCTCAAGCTCGGCTGACCAACTACCGCGCGCTACCGGACT 312  
QY 379 CAGATCACCTGTGGCAGGCGCCCTTGTGTGAGATCAAGGTGGCGGCGCAGATCAAGGAG 438  
Db 313 GGCTTGTGTGCGCGCGCTGTGCTGAAGAAGCTCAACCTCGACTTCAAAGTACGAGGGT 372  
QY 439 GCCTCTGTGACACCGCGCGCAGACACCGCTCTGAGGAGATGAGCCTGCCCGGCAAG 498  
Db 373 GTCAAGAAGGTCAACGGCGAGGACTAACGTGAGAGCTCGACGAGGCGCCCGCGCT 432  
QY 499 TGAAGCCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCCAGTACGACAG 558  
Db 433 TTCAAGGCCCTGTCTCGACGCTCGGCTGGTCCGCACTCGACTGGCGCGCGCTTTCGCC 492  
QY 559 ATCTGTATGAGATCTTGGCGCAAGAGGCCATCGGACCGGTGTGATCGGCGCCACCCCC 618  
Db 493 GCCTCAAGGGCATGTGCGACGCGCGCGTCAACGTCCCC---CACAGCGAGACCGCGTTC 549  
QY 619 GTGAATCATCTCGCGCGCAACATGCTGACCCAGCTGGGCTGCACTTGAATTTCCCCATC 678  
Db 550 GTCGGCTTCAACGGCGACAAAGAAGAGGACTCAAGCGCGGCTCTCCCAAGTATCATCTTC 609  
QY 679 AGCCCCATCGAGACCGTGGCCCGGTGAAGCTGAAGCCCGGCAATGGACGGCCCCCAAGGTGAAG 738  
Db 610 GGCGGCCACGTGCGCGGTACATGAAGCTCTCAAGGAGCAGGACGCGCGCGCTTCGAC 669  
QY 739 CAGTGGCCCTGACCGGAGGAGAATCAAGGCCCTGACCCCATCTTCGCGAGGAGATGAG 798  
Db 670 CGCCAGTTCTCGCGCTACGCCAAGAGGGGTGTCCCGCCGACATGCTCGAGAGAGATCTAC 729  
QY 799 AAGGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTTACAACACCCCGCTTTCGCC 858  
Db 730 ACCGAGGCCCAAGCAGAGATCCCGCGCGGACCCGACCTTCGTCCCCAAGCGCGCTCGAAG 789  
QY 859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTTCGCGGAGCTTGAACAAG 918

Db 790 CCCGAGGGCGCCACGCCCAAGCCTTGGGCAAGCGCAGGCTGACGTACCACGAGCGCAAG 849

Qy 919 CGCACCCAGGACTTCTGGAGAGGTGTCAGCTGGGCATCCCCCACCCTGGCC 966

Db 850 AACCGCGTCGCCCAAGAAGAGGTCCGCTGGGCTACCCGACGCCCCC 897

RESULT 12  
 BM320900 1165 bp mRNA linear EST 03-JAN-2000  
 LOCUS rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library  
 DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.

ACCESSION BM320900  
 VERSION BM320900.1 GI:18055306  
 KEYWORDS EST.  
 SOURCE Mastigamoeba balamuthi  
 ORGANISM Mastigamoeba balamuthi  
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.

REFERENCE 1 (bases 1 to 1165)  
 AUTHORS Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Senses,C.W., Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and Philippe,H.  
 TITLE The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
 MEDLINE 21819461  
 PUBMED 11830664

COMMENT Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockvax.rockefeller.edu  
 Insert Length: 1165 Std Error: 0.00  
 POLYA=No.

FEATURES  
 source Location/Qualifiers  
 1..1165  
 /organism="Mastigamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
 /note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match 3.2%; Score 79.6; DB 4; Length 1165;  
 Best Local Similarity 45.1%; Pred. No. 0.00039;  
 Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3

Qy 90 CATCAAGTGTCTTAACTGCGGCAAGGAGGGCCACATCGCCGCAACTGCGCGCCCCCG 149

Db 46 CGTCAAGAACGAGGGTACTTCAAGCGGTTCCAGACCAGTTCCGTGCGCGCGCGAGG 105

Qy 150 CAAGAAGGGCTGTCTGGAGTGTGCGGCAAGGAGGGCCACCAGATGAAGGACTGCACCGAGC 209

Db 106 CAAGACGGACTACCGCGCGCGCCANCCTGGTGATCCAGGACAAGAACAGTACACAGCCC 165

Qy 210 CGAGGCCAACTTCTTTCCGCGAGGACCTGGCCCTTCCCCCAGGGCAAGGCCGCGAGTTCCC 269

Db 166 CAAGTACCGCTTCGTGTC-----GTCCCGCTTACCACAGGGGACATCGTCTGCAGATCGC 219

Qy 270 CAGCGACGAAACCGCGCCCAACAGCCCCACCAGCGCGGAGCTGCAGGTGCGGGGCGCA 329

Db 220 CTACGCCAAGATCGACGCGGACCAATCCTCGCGCGCGCCCTACTCGCACGAGCTCACCG 279

Qy 330 CCCC CGCAGCGAGCGCGCGCGCGGACCGGACCCCTGAACCTTCCCCCAGATCACCCCT 389

Db 280 CTTGCGGTCAAGCTCGGCTGACCACTACGCCCGCGCTTACCGCACTGCGCTGCTGCT 339

Qy 390 GTGCGACGCGCCCCCTGTGTGAGCATCAAGGTTGGGGGGCGAGATCAAGAGGAGGCCCTGCTGGA 449

Db 340 GGCCCGCGGTGTGCTGAAGAAGCTCAACCTCTACTTCCAAAGTACGAGGCTGTCAAGAGGT 399



```

/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Wastigamoeba balamuthi lambda ZAP II Library"
/notes="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match      3.2%; Score 79; DB 4; Length 867;
Best Local Similarity 45.7%; Pred. No. 0.00049;
Matches 314; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

QY 58 CACGCGAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGGAG 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 CCGCACAAGATGCGCGAGTGCCTCGCGGTTCATCATCTGCTGCGCAACAGGTTGAAGTAC 190

QY 118 GGCACATCGCCGCAACTCGCGGCCCCCGGCAAGAGGGCTGCTGGAAGTGGCGCAAG 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 GCCTGACCGCGTGAAGTCACTCGATCGTGAAGCGCTGATCAAGATCGACGGC 250

QY 178 GAGGCGCACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTG 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 AAGTTCGCACACGACGCACTTCCCGGGCTTCATGACGCTGCTCTCGATCGACAAG 310

QY 238 GCCTTCCCCCAGGCGAGGCCCGCGAGTTTCCCAGCGAGCAGAAACCGCGCCAAAGCCCC 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 ACCGACGAGCACTTCCGCTGTCTACGACACCAAGAGGCGCTTCCAGGCGCACCGCATC 370

QY 298 ACCAGCGCGAGCTGACGTT--GCGGGCGACAAACCCCGCAGCGAGGCGCGGCGGAG 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 AACTCGACGAGGCCAAGTTTCAAGTCTGGCAAGGTCGCGCGCTGCGAGCTCGAGTCAACAAG 430

QY 355 CGCAGGGGCACCTGAACTTCCCGCCAGATCACCTGTGGCAGCGCCCGCTGTGGAGCATC 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 GGCATCCCGTACTGTGTGACCCGACGCGCGGACAGATCGCTACCCCAACCCCGACATC 490

QY 415 AAGTGGGGCGGCAGATCAAGGAGGCGCTGCTGGAACACCGCGCGCGACACGACCTGTG 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 AAGGTCAACGACACGCTCAAGATCGACCTGCGCTCGGGCAAGATCATCGACTTCGTGAGG 550

QY 475 GAGGAGTGAAGCTGCGCGCGAGTGGAGAGCCCAAGATGATCGCGGCATCGCGGCTTC 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 551 TTCGAGATCGGCAACTCTGTCATGATCACTGGCGGAGCGCAACCTTGGCGCGCTCGGCGTC 610

QY 535 ATCAAGTGGCGCAGTACGACACAGATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGC 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 611 ATTGTGGCGCGCAGAGACGAGGGCTGCTTGAGATCATTCACGTCAGAGGACGCGTC 670

QY 595 ACCGTGCTGATCGGCGCCACCCCGGTGAACATCATCGGCGCGCAACATGTGACCCAGCTG 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 671 GGCACACAGTTTCGCGACGCGCTTGAACCAAGCTTCTGATGATCGCAAGGCGACCAAGTCC 730

QY 655 GGTGACACCTGAACCTTCCCATCAGCCCATCGAGACCGTGCCTGAGCTGAAGCCC 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 731 CTCGTACGCTGCGCGCGGCGGATCAAGAGTTCGATCATCGAGGAGTTCAGGGCG 790

QY 715 GGCATGAGCGGCCCCCAAGGTGAAGCAG 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 791 CGCACCGCACAGGACCGAGGAGGAG 817

RESULT 14
LOCUS CL962901
DEFINITION OsiFC08219 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL962901
VERSION CL962901.1
KEYWORDS GI:52380530
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Ehrhartoideae; Oryzeae; Oryza.
```

```

1 (bases 1 to 11691)
Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,Z., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
Location/Qualifiers
    1..11691
    /organism="Oryza sativa (indica cultivar-group)"
    /mol_type="genomic DNA"
    /db_xref="taxon:39946"
    /clone_lib="Oryza sativa Express Library"
    /note="Oryza sativa exon trapped genomic sequences"

ORIGIN
Query Match      3.2%; Score 78.4; DB 9; Length 11691;
Best Local Similarity 42.2%; Pred. No. 0.00074;
Matches 567; Conservative 0; Mismatches 771; Indels 6; Gaps 2;

QY 150 CAAGAAGGGCTGTGGAAGTGGCGCAAGAGGGCCACAGATGAAGACTGCACCGAGCG 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2991 CAAGAAGGGCTTTCACAGCAGGCGCTGAAGTGCAGAGTGGGCGTGAAGATCGCGGAAGT 3050

QY 210 CCAGGCGCAACTTCTCCGCGAGGACCTGGCTTCCCGCAGGGCAAGCGCCGAGTTTCCC 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3051 GAAGAAGGGCGGCGACGCGGTGACCGTGGCTTACCGATGCCAGGGCGAGCGCCGCGA 3110

QY 270 CAGCGAGCAGAAACCGCGCCAAACAGCCCCCAGCGCGAGCTGCAGGTGCGCGCGCAAA 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3111 ACTGGCGGTGGAACAAGCTCATCGTCTCCATCGCGCGCGTGCACCAACACCATCGCGCTCAA 3170

QY 330 CCCCAGGAGGAGCGCGCGCCAGCGCCAGAGGCGACCTGGAATCTCCCGCAGATCACCT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3171 CCCGGAAGCGTGGGCGCTGAAGCTCAGCAGCGCGCGCCATCGTGGTCGATCGCGAGCTG 3230

QY 390 GTGCGAGCGCCCTCTGTGAGCATCAAGTGGCGCGCAGATCAAGAGAGGCGCTGCTGGC 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3231 CAAGAACAACCTGCCCCGCGTGTGGCGGTGGGCGAGCTGTGTGCGGCGCCGATGCTGGC 3290

QY 449 --ACACGCGCGCGACACACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAGCC 506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3291 GCACAGGCGGAGAGAGGCGCTGCGGTGGCGGAGCGCATCGCGGCGAGCATGGCCA 3350

QY 507 CAAGATGATCGGGCGGATCGCGGCTTCATCAAGGTGCGCGAGTACGACAGATCTCTGAT 566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3351 CGTGAACCTTCAACACCATCCCTTGGGTTCATCTACACAGCCCCGAGATCCGCTGGGTGGG 3410

QY 567 CGAGATCTCGGCAAG---AAGGCCATCGGCACCGTGTCTCGGCCCGCCACCCCGCTGAA 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3411 CCGCACGAGCAGCAGCTCAAGGCCGACGCGGTGCGAGTACAAGCGCGGACCTTCCCGTT 3470

QY 624 CATCATCGGCGCGCAACATGTGACCCAGCTGGGCTGCACCTTGAACCTTCCCGCATCAGCCC 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3471 CCTGGCGAAGCGCGCGCGCTGCGGTGGCGGACACAGCGGCATGTCTCAAGTCTCTGGC 3530

QY 684 CATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGACCGGCCCAAGGTGAAGCATG 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3531 CGATGCGGCCACGAGAGATCTCTGGGCGGTGCATCTGTGGGCGCGCATGTGCGCGAGCGCT 3590

QY 744 GCCCTGACCGGAGGAGATCAAGGCCCTGACCGGCATCTGCGAGGAGATGAGAGGA 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3591 GATCTCGAGGCGCGTGTGGCCATGAGTTCAAGGCCAGCGCGGAGGACATCGCGCGCAT 3650
```

804 GGGCAAGATCACCAAGATCGCCCGCGAGAACCCCTACAAACCCCGGTGTTCCGCATCA 863  
 Db |||||  
 3651 CTGCCACGCCCATCCCTCGCTTTCCGAGCGACCAAGGAAGCGCCCTGGCGCGTATGA 3710  
 Qy |||||  
 864 GAAGAAGGACAGACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCAC 923  
 Db |||||  
 3711 GGCCGAGCTGCGCGCAAGGGTTCCGACGACCCCGGCCAGCTGCGCGCGCTCGAGGC 3770  
 Qy |||||  
 924 CCAGGACTTCTGGGAGTGGAGTGGGATCCCCACCCCGCGCTGGAAGAAGAA 983  
 Db |||||  
 3771 GCTGAGCGCTGCGCGGACGACTGGCGCGCTACAAAGAAAAGCGCTCAACGCGCTCAA 3830  
 Qy |||||  
 984 GAGCGTGACCTGTGAGCTGGCGAGCGCTACTTACGCTGGCCCTTGGACGAGGACTT 1043  
 Db |||||  
 3831 GAAGCTCATCCCGCCCGAGGTGGCGCGGCTGTACATGTACGGGGCGTGGGCG 3890  
 Qy |||||  
 1044 CCGCAAGTACACCGCTTACCATCCCGAGATCAAAACAGAGACCCCGGCATCCGCTA 1103  
 Db |||||  
 3891 TGGCAAGAGTCTCTGATGAGTCTCTTCAACGCGGTGCGATCGCCGCAAGGTGG 3950  
 Qy |||||  
 1104 CCAGTACAAGCTGCTGCCCGAGGCTGGAAGGGCAGCCCGCATCTTCCAGAGCAGCAT 1163  
 Db |||||  
 3951 GCTGCAATTCACGAATTCATGCTGAGTGCACCGGAACTGGCGGCGCTGCAGGGCAC 4010  
 Qy |||||  
 1164 GACCAAGATCTGGAGCCCTTCCGGCGCGCAACCCCGAGATCTGTATCTACAGTACAT 1223  
 Db |||||  
 4011 CGTGAACCCGCTGAGCTGCTGGCGAGCGCATCGCCAAAGCGCTTACAGCTCATCTGCTT 4070  
 Qy |||||  
 1224 GGAGGACTGTACCTGGCGAGCGACTGGAGATCGGCCAGCACCGCGCAAGATCGAGGA 1283  
 Db |||||  
 4071 CGAGCAATTCAGTGGCGGACATCAAGGAGCCATGATCTGACCGCTGTGGCGGC 4130  
 Qy |||||  
 1284 GTGCGCAAGACCTGTGGCTGGGCTTCCACACCCCGCAAGAGCAACCAAGAGGA 1343  
 Db |||||  
 4131 GCTGTTGCAACCGCGTGGCTTCTGTCACCACTCGAACTTCAAGCCGACGACCTGTA 4190  
 Qy |||||  
 1344 GCGCCCTTCTGTTGGATGGCTAGAGCTGCACCCCGCAAGTGGACCGTGGAGCCCAT 1403  
 Db |||||  
 4191 CCGGGGGCGCTGCACCGGACCGCATCTCTCGCGCCATCGATCTGCTCAACGAGAAGCT 4250  
 Qy |||||  
 1404 CGAGTCCCGGAGAGAGAGCTGAGCGTGAACGACATCCAGAACTGGTGGGCAAGCT 1463  
 Db |||||  
 4251 GGAGTGGTGAACGTGAGCAACGACCGACTACCGGGCGCGACGCTGGAGCAGGTCCG 4310  
 Qy |||||  
 1464 GAATGGGGCCAGCGATCTACCC 1487  
 Db |||||  
 4311 GCTCTACCATGCGCGTGGGCCC 4334

## RESULT 15

CL961326 1962 bp DNA linear GSS 21-SEP-2004  
 LOCUS OaIFC005994 Oryza sativa Express Library Oryza sativa (indica  
 DEFINITION cultivar-group) genomic, genomic survey sequence.

CL961326  
 VERSION CL961326.1 GI:52377391

KEYWORDS GSS.

SOURCE Oryza sativa (indica cultivar-group)

ORGANISM Oryza sativa (indica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 1962)

Jiao, Y., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,

Wong, G.K.S., Deng, X.W. and Wang, J.

An analysis of transcriptional regulation of the rice genome and

its comparison to Arabidopsis

Unpublished (2004)

CONTACT: Chen Chen

Department of Bioinformatic

Beijing Institute of Genomics

Chinese Academy of Sciences, Beijing 101300, China

COMMENT

Tel: 86-10-80481559  
 Fax: 86-10-80488676  
 Email: chenchen@genomics.org.cn  
 Rice genomic sequence.  
 Class: exon-trapped.  
 Location/Qualifiers  
 1..1962  
 /organism="Oryza sativa (indica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3946"  
 /clone\_lib="Oryza sativa Express Library"  
 /note="Oryza sativa exon trapped genomic sequences"

## FEATURES

source

## ORIGIN

Query Match 3.2%; Score 78.2; DB 9; Length 1962;  
 Best Local Similarity 41.6%; Pred. No. 0.00071;  
 Matches 737; Conservative 0; Mismatches 1013; Indels 21; Gaps 3;  
 Qy 291 CAGCCCCACCAGCCGCGAGCTGAGGTGCGGGCGACAAACCCCGCAGCAGGCGCGGC 350  
 Db |||||  
 162 CATCGCGACGCCGCCAAGAACCCAGGTGCGCATGAACCCCAACACACCGCTTTTGATGC 221  
 Qy |||||  
 351 CGAGCGCAGGGCACCTGAACTTCCCCAGATCACCTGTGGCAGCGCCCTTGGTGAG 410  
 Db |||||  
 222 AAGCGGCTGATCGGGCGGCGCTTCTCGACCCCTCGTGCAGCGGACATGAAGATGTG 281  
 Qy |||||  
 411 CATCAAGTGGGGCGGCAGATCAAGAGGCGCTCTCGACACCGCGCGCCAGCACCGT 470  
 Db |||||  
 282 GCCGTTCAAGTGTGCTGCGCCAGCCGCAAGCCGATGATCTGTTGAGCTACAAGGG 341  
 Qy |||||  
 471 GCTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGGCGCATCGGGCG 530  
 Db |||||  
 342 CGAGGAGAAGAAGTTCTCGCGGAGGAGATCTCGTTCATGTTGCTCAACAAAGTGAAGGA 401  
 Qy |||||  
 531 CTTTCATCAAGTGGCGCAGTACGACGATCTCTCATCGAGATCTCGGCAAGAAGGCCAT 590  
 Db |||||  
 402 GATCGCGAGGCTTCTCTCAGCACCAACATCAAGAACGCGCTCATACCGTCCCGGCTA 461  
 Qy |||||  
 591 CGGACCGCTGTGATCGGCGCCCAACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCA 650  
 Db |||||  
 462 CTTCAAGCTCCCGAGCGCCAGGCCACCAAGGAGCGCGGCTCATCTCCGCGCTCAACGT 521  
 Qy |||||  
 651 GCTGGGTGACCTGAACTTCCCATCAGCCCATCGAGCCGCTGAGAGCTGAA 710  
 Db |||||  
 522 CATGCGCATCATCAACGAGGCCACCGCGCGCCATCGCTTACGCGCTCGCAAGAAGGC 581  
 Qy |||||  
 711 GCCCGCATGAGCGGCCCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 770  
 Db |||||  
 582 CGCCAGCACCGCGGAGAAAGTGTCTCATCTTCGACCTCGGCGCGGCGACCTTCGACGT 641  
 Qy |||||  
 771 CCTGACCGCATCTGCGAGGAGATGAGAGAGGGGCAAGATCAACAGATCGGCCCC - 828  
 Db |||||  
 642 CTCCATCTCACCATCGAGGAGGGCATCTTCGAGTCAAGGCCACCGCGCGACACCA 701  
 Qy |||||  
 829 -----GAGAACCCCTACAAACCCCGTGTTCGCATCAAGAGAGAGAGAGAGAGAGAGAG 881  
 Db |||||  
 702 CTTGGAGGCGAGGACTTCGACAAACCGGATGGTGAACCACTTCTGTCGAGGAGTTCAGAG 761  
 Qy |||||  
 882 GTGCGCAAGCTGTGGACTTCGCGAGCTTTCGCGAGCTGMAACAAGCGCACCCAGGACTT -CTG 935  
 Db |||||  
 762 GAAGCAAGAAGAGACATACCGGAAACCGAGGCGCTTCGCGCGCTGAGGAGCGGCTG 821  
 Qy |||||  
 936 GAGAGTGCAGCTGGGCATCCCCCACCCCGCGCTTGAAGAAGAAGAGAGAGAGAGAGAGAG 995  
 Db |||||  
 822 CGAGGCGCAAGAGGAGCGCTGCTCTCAACCGGCCAGACCAACCATCGAGATCGACTCGCT 881  
 Qy |||||  
 996 GCTGAGCTGGGCGACCGCTTCTTACGCTGCCCTCTGGAAGAGAGAGAGAGAGAGAGAGAG 1055  
 Db |||||  
 882 CTACGAGGCGATCGACTTCTACGCGACCATCACCGCGCGCGCTTCGAGGAGCTCAACAT 941  
 Qy |||||  
 1056 CGCCTTTCACCATCCAGGATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1115  
 Db |||||  
 942 GGACCTTCTCCGCGGTGTCAGGAGCGCGTGAAGAGTGTCTCGCGAGAGAGAGAGAGAGAG 1001

```

QY 1116 GCTGCCCCAGGGCTGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1002 CAAGGCCAGATCCAGATGCTGCTCTCTGTCGGAGGCTCCACCGTATCCCCAAGGTGA 1061
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1176 GGAAGCCCTTCCGGCCCGCAACCCCGAGATCGTGAATACAGTACATGAGACGACCTGTA 1235
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1062 GCAGCTCTGCAGGACTTCTTCAACGGGAAGGAGCTCTGCAAGAGCATCAACCCCGACGA 1121
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1236 CFTGGGCAGCGACTGAGAGTCGGCAGCAGCCGGCCAGATCGAGGAGCTGCGCAAGCA 1295
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1122 GGCCTGTGGCTACGGGCGCCGCTCCAGGCGCCCATCTCAGCGGAGGGGCAACCAAG 1181
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1296 CTTGCTGCGCTGGGCTTACCAACCCCGCAACAGAGCAGAGGAGCCCCCTTCTCT 1355
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1182 GGTGCAGGACTGCTCTCTCTGAGCTCACCGCTCTCGCTCGGTTGGAGAGGCGCG 1241
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1356 GTGGATGGGTACGAGCTGACCCCGCAAAAGTGGAGCGCTGAGCCCATCGAGCTGCCCGA 1415
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1242 GGGTGTCAACCGTGTGATCCGAGGAACACCAACCTCCCAACCAAGAAGGAGCAGGT 1301
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1416 GAAGGAGAGCTGACCGTGAACACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCCAG 1475
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1302 CTTCTCACCTACTCCGACAAACAGCCCGCGTCTCATCCAGGTGTACGAGGGCGAGAG 1361
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1476 CCAGATCTACCCCGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCCAAGGC 1535
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1362 GACGAGGACCAAGGACNACAACTGTCTGCGAAGTTGAGCTCACCGGCATCCGCGCGC 1421
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1536 CTTGACCGCATCGTGCCCTGACCGAGAGGCGGAGCTGGAGCTGGCCGAGAACCCCGA 1595
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1422 GCCGAGGGGGTGGCCCGAGATCAACGTGACGTTTCAACATCGACGCGAAACGGCATCTGAA 1481
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1596 GATCTCTGGCGAGCCGTGCAAGCGGTGTACTAGCA-----CCCCAGCAAGGACCTGGT 1649
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1482 CGTGTGCGCGGAGGACAAAGACGACGCGGGAAGAAAGAAACAGATCACCATCAACCAACGCAA 1541
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1650 GGCGGAGATCCAGAAGCAGGGCCACACACAGTGTGACCTTACAGATCTTACCAGGAGCCCTT 1709
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1542 GGGCGGCTGAGCAAGGAGAGATCGAGCGGAGTGGTGCAGGAGGCGGAGAGTACAAGGC 1601
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1710 CAAGAACCTGAAGACCGGCAAGTAGCGCAAGATGCGCACCGCCCAACCAACCAACGCTGAA 1769
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1602 CGAGGACGAGCAGTGTGCGGCAAGGTGGAGGCGCGCAACGCGCTGGAGAACTACGCGTA 1661
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1770 GCAGCTGACCGAGCCGTGCAAGATCGCCATGAGAGCATCGTGTATCTGGGGCAGAC 1829
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1662 CAACATGCGCAACACGCTGCGCGACGAGAAAGATGCGCTTCAAGCTCCCCCGCCGACGACAA 1721
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1830 CCCCAAGTTCGGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTG 1889
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1722 GAAGAGATCGAGGACGCCATCGAGGATGCCATCAAGTGTCTGACGGCAACCAAGCTGCG 1781
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1890 GCAGGCCACTGTGATCCCGAGTGGAGTTCGTGAACACCCCCCTCTGTGAAGCTGTG 1949
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1782 CGAGGCGGACGAGTTCGAGGACAAAGATGAAGGAGCTCGAGAGCTCTGCAACCCGATCAT 1841
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1950 GTACCAGCTGAGAGGAGCCCATCATCGCGCGGAGACCTTCTACGTGAGCGGCGCGC 2009
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1842 CTCAAAGATGTACAGGGCGGCGCGGTGTGTCGGGGGATGAGCAAGACGCCGCCAA 1901
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 2010 CAACCGGAGACCAAGATCGCAAGGCGCGC 2040
Db      |||||      |||||      |||||      |||||      |||||      |||||
QY 1902 CGGGGTGCCGGCACCGGCGGTGGCAGCGC 1932
Db      |||||      |||||      |||||      |||||      |||||      |||||

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 03:13:23 ; Search time 34 Seconds  
(without alignments)  
10815.332 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 4573

Sequence: 1 gtcagcgcaccatgcccga.....gggctagcaccgtgaattc 2463

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09610313/runat\_31052005\_155137\_15173/app\_query.fasta\_1.7893  
-DB=Issued Patents AA -QPMF=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CGN 1 1 101 @runat\_31052005\_155137\_15173 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3497	76.5	1005	2	US-07-743-357-1
2	3493.5	76.4	1003	2	US-07-743-357-9
3	3475	76.0	1015	3	US-08-463-210-9
4	3475	76.0	1015	3	US-09-124-900-3
5	3475	76.0	1015	4	US-08-463-028-9
6	3475	76.0	1016	2	US-07-743-357-2
7	3472	75.9	1003	4	US-09-309-572-17
8	3472	75.9	1003	4	US-09-718-096-17
9	3469	75.9	1016	2	US-07-743-357-5
10	3468	75.8	1016	2	US-07-743-357-3
11	3465	75.8	1016	2	US-07-743-357-4
12	3459.5	75.7	1003	2	US-07-743-357-10

13 3456.5 75.6 1003 2 US-07-743-357-8  
14 3449 75.4 1004 2 US-07-743-357-7  
15 3183.5 69.6 1014 4 US-09-319-588C-6  
16 3141 68.7 913 2 US-07-743-357-22  
17 3119.5 68.2 1350 4 US-09-952-060-35  
18 3078 67.3 913 2 US-07-743-357-6  
19 2822.5 61.7 875 4 US-09-952-060-6  
20 2817 61.6 850 4 US-09-952-060-2  
21 2798.5 61.2 875 4 US-09-952-060-8  
22 2793 61.1 850 4 US-09-952-060-4  
23 2789 61.0 562 3 US-09-117-217-14  
24 2789 61.0 562 3 US-09-735-487-14  
25 2758 60.3 560 4 US-10-205-641-1  
26 2507.5 54.8 1018 4 US-09-206-551-46  
27 2483.5 54.3 1016 4 US-09-625-972-24  
28 2259 49.4 441 4 US-09-991-258-16  
29 2212 48.4 1055 2 US-08-659-251-5  
30 2212 48.4 1055 3 US-09-256-490-5  
31 2212 48.4 1055 5 PCT-US96-11445-5  
32 2193.5 48.0 3080 6 5223423-4  
33 2193.5 48.0 3080 6 5223423-4  
34 2154 47.1 427 3 US-09-690-265-1  
35 1572 34.4 327 3 US-08-679-493A-72  
36 1555 34.0 327 3 US-08-679-493A-71  
37 1552 33.9 314 1 US-08-589-446-6  
38 1552 33.9 314 1 US-08-444-882-6  
39 1552 33.9 314 2 US-08-389-459A-6  
40 1552 33.9 314 3 US-08-987-967A-6  
41 1311 28.7 1150 3 US-09-238-303-9  
42 1311 28.7 1150 4 US-09-946-239-9  
43 1309 28.6 275 6 5252477-3  
44 1309 28.6 275 6 5252477-3  
45 1297 28.4 261 6 5320958-3

#### ALIGNMENTS

#### RESULT 1

US-07-743-357-1

; Sequence 1, Application US/07743357

; Patent No. 5858646

; GENERAL INFORMATION:

; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological

; NUMBER OF INVENTION: activity for use as diagnostic reagent and/or vaccine

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: KIRBY EADES GALE BAKER

; STREET: Box 3432, Station D

; CITY: Ottawa

; STATE: Ontario

; COUNTRY: Canada

; ZIP: K1M 1H8

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/743,357

; FILING DATE: 21-AUG-1991

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/CA90/00062

; FILING DATE: 23-FEB-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Gale, Edwin J.

; REGISTRATION NUMBER: 28,584

; REFERENCE/DOCKET NUMBER: 30924-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (613) 237-6900

; TELEFAX: (613) 237-0045

; INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1005 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
; US-07-743-357-1

Alignment Scores:
Pred. No.: 1,22e-208 Length: 1005
Score: 3497.00 Matches: 655
Percent Similarity: 95.56% Conservative: 34
Best Local Similarity: 90.85% Mismatches: 26
Query Match: 76.47% Indels: 6
DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x US-07-743-357-1 (1-1005)
QY 220 TTCTTCGCGAGGACCTGGCTTCCCGCCAGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279
DB 2 PhePheargGluAspLeuAlaPheLeuGlnGlyAlaArgGluPheSerSerGluGln 21
QY 280 AACCGGCGCAACGCCCCACAGCGCGAGCTGAGGTG-----CGCGCGCAACACCCC 333
DB 22 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 41
QY 334 CGCAGCGAGGCGCGCGCGAGCCGAGCCGAGCCACCTG-----AACTTCCCGCCAGATCAC 387
DB 42 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 61
QY 388 CTGTGGCAGCGCCCTGGTGCAGCATCAAGTGGCGCGCGCGAGATCAAGGAGGCGCTGCTG 447
DB 62 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 81
QY 448 GACACGCGCGCGAGCACCGTGTGGAGAGATGAGCTGCCCGCGCAAGTGGAGGCC 507
DB 82 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 101
QY 508 AGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCGCGAGTACGACGACGATCCTGATC 567
DB 102 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIle 121
QY 568 GAGATCTGCGCGCAAGAGGCGCATCGGCACCGTGTGATCGCGCGCGCGCGCGCGCGCG 627
DB 122 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 141
QY 628 ATCGCGCGCAACATGTGACCCAGCTGGTGGCTGACCCCTGAACCTTCCCGCATCAGCCCATC 687
DB 142 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 161
QY 688 GAGACGCTGCGCTGAGCTGAGCCCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
DB 162 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 181
QY 748 CTGACCGGAGAGAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAGAGAGAGGCG 807
DB 182 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 201
QY 808 AGATACCAAGATCGCGCGCGAGAACCCCTACAACACCCCGCTGTGTCGCATCAAGAAG 867
DB 202 LysIleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys 221
QY 868 AAGCAGACCAAGTGGCGCGAGCTGCTGCTCCGCGAGCTGACACAGCCGCGCGCGCG 927
DB 222 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgGln 241
QY 928 GACTTCTGGGAGGTGACGCTGGGCATCCCCCACCCTGCGCGCGCGCGCGCGCGCGCG 987

```

```

DB 242 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 261
QY 988 GTGACCGTCTGCGAGCGCGCGCTACTTTCAGCGTCCCGCGCGCGCGCGCGCGCGCG 1047
DB 262 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArg 281
QY 1048 AAGTACACGCGCTTACCATCCCGAGCATCAACACGAGAGACCCCGCGCATCCCGCTACCAG 1107
DB 282 LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln 301
QY 1108 TACAACGTGCTGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCAGAGCAGCATGACC 1167
DB 302 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 321
QY 1168 AAGATCTGGAGCCCTTCCGCGCGCGCGCGAGATCGTGATCTACCAG-----GCC 1221
DB 322 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTrpGlnTrpMetAsp 341
QY 1222 CCCTGTAGCTGGCGAGCGACCTGGAGATCGCGCGAGCAGCCCGCGCGAGATCGAGAGCTG 1281
DB 342 AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 361
QY 1282 CGCAAGCAGCTGTGCTGGGCTTTCACACCCCGCAAGAACGACCAAGAGGAGCC 1341
DB 362 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 381
QY 1342 CCCTTCTGTGGTGGCTACGAGCTGCGCGAGTCCCGCGAGTGGACCTGGAGCCCGCATCGAG 1401
DB 382 ProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal 401
QY 1402 CTGCCCGAAGAGAGAGCTGGACCGCTGAAACGACATCCAGAAGCTGGTGGCGAAGCTGAAC 1461
DB 402 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 421
QY 1462 TGGCGCAGCCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGAAGCTGTGCGCGCG 1521
DB 422 TrpAlaSerGlnIleTrpProGlyIleLysValArgGlnLeuLysLysLeuLeuArgGly 441
QY 1522 GCCAAGGCGCTGACCGACATCGTGGCGCGCGCGAGGCGCGAGCTGGAGCTGGCGCGAG 1581
DB 442 ThrLysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGlu 461
QY 1582 AACCGCGAGATCTCGCGCGAGCGCGCTGTACTACCGACCCCGAGCAAGAGGAGCTG 1641
DB 462 AsnArgGluIleLysGluProValHisGlyValTrpAspProSerLysAspLeu 481
QY 1642 GTGCGCGAGATCCAGAAGCGAGGCGCGAGCTGTGAGCTGTACAGATCTACAGAGGAGCC 1701
DB 482 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTrpGlnIleTrpGlnGluPro 501
QY 1702 TTCAAGAACTGAGACCGCGAGTACGCGAGTACGCGAGCGCGCGCGCGCGCGCGCGCG 1761
DB 502 PheLysAsnLeuLysThrGlyLysTrpAlaArgMetArgGlyAlaHisThrAsnAspVal 521
QY 1762 AAGCAGCTGACCGAGGCGCGTGGCAAGAGATCGCCATCGAGAGCATCGTGATCTGGGCGAAG 1821
DB 522 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 541
QY 1822 ACCCCCAAGTTCGCTGCGCGCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACTAC 1881
DB 542 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpGluTrp 561
QY 1882 TGGCAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCGCGCGCGCGCGCG 1941
DB 562 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 581
QY 1942 TGGTACAGCTGGAGAGAGGCGCGCATCATCGCGCGCGCGAGACCTTCTACGTGGAGCGCG 2001
DB 582 TrpTrpGlnLeuGluLysGluProIleValGlyAlaGluThrPheTrpValAspGlyAla 601
QY 2002 GCCAAGCGGAGACCAAGATCGGCAAGCGCGCTGTGATGACCGAGCGCGCGCGCGCGAG 2061
DB 602 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTrpValThrAsnArgGlyArgGlnLys 621

```





Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 QY 1171 ATCTGTGAGCCCTCCGCGCCCGACACCCGAGATCGTGTATCTACAG-----GCCCCC 1224  
 Db 321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
 QY 1225 CTGTACCTGGCGCAGCAGCTCGAGATCGCGCAGCAGCCGCCCAAGATCGAGAGCTCGCGC 1284  
 Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360  
 QY 1285 AAGCACTGTGCGCTGGCGCTTACACACCCCGCAGCAAGACCAAGAGAGAGCCGCC 1344  
 Db 361 GluHisLeuLeuLysTyrGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
 QY 1345 TTCTGTGGATGGCTACGAGCTGCACCCCGACAGTGGAGCCGTGAGCCCATCGAGCTG 1404  
 Db 381 PheLeuTyrMetGlyTyrGluLeuHisProAspLysTyrThrValGlnProIleGlnLeu 400  
 QY 1405 CCCGAGAAGAGAGCTGGACCGCTGAACGACATCCAGAGCTGCTGGCGAAGCTGAACCTGG 1464  
 Db 401 ProAspLysGluSerThrThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTyr 420  
 QY 1465 GCCAGCCAGATCTACCCCGCGCATCAAGTGGCGCAGCTGTGCAAGTGTGCGCGCGCC 1524  
 Db 421 AlaSerGlnIleTyrProGlyIleLysValLysValLysGlnLysLeuLeuArgGlyAla 440  
 QY 1525 AAGCCCTGACCATCTGCGTGGCCCTGACCGAGAGCCGAGCTGGAGCTGGCGCGAGAC 1584  
 Db 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460  
 QY 1585 CGCGAGATCTCGCGAGCCGCGTGCACGCGTGTACTACCAACCCCGCAGAGACCTGGTG 1644  
 Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480  
 QY 1645 GCCGAGATCCAGAGAGCGCCACGACCTAGTGGACCTTACAGATCTACAGAGACCTTC 1704  
 Db 481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluGlnTyr 500  
 QY 1705 AAGAACTGAGACCGCAGTACGCAAGATGGCAGATGGCAGCCGCCACACACGACGCTGAG 1764  
 Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
 QY 1765 CAGCTGACCGAGCGCGTGCAGAGATCGCATCGAGAGCATCTGTATCTGGCGCAAGACC 1824  
 Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnLysSerIleValIleTyrGlyLysThr 540  
 QY 1825 CCCAAGTTCGCTGCCATCCAGAGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884  
 Db 541 ProlLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTyrThrGluTyrTrp 560  
 QY 1885 CAGGCCACTGGATCCCGAGTGGAGTTCGTGACACACCCCGCTGGTGAAGCTGGG 1944  
 Db 561 GlnAlaThrTrpIleProGluTyrGluPheValAsnThrProProLeuValLysLeuTyr 580  
 QY 1945 TACGAGCTGAGAGAGGAGCCATCTCGCGCCGAGACCTTCTACGTGGACGCGCGCCG 2004  
 Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 QY 2005 AACCGCGAGACCAAGATCGCAAGCGCGGTACTGTACCGACCGCGGCGCGCAGAGATC 2064  
 Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
 QY 2065 GTGAGCTGACCGAGACCAACCAAGACGAGCTGCAGGCCATCCAGCTGGCCCTG 2124  
 Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
 QY 2125 CAGGACACCGCGCAGAGTGAACATCTGACCGACAGCCAGTACGCGCTGGGCATCATC 2184  
 Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 QY 2185 CAGGCCAGCCGCAAGAGCGAGAGCTGGTGAACCCAGATCATCGAGAGCTGATC 2244  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

QY 2245 AAGAAGCAGAGGTGTACCTGAGCTGGTGGCCGCCCAAGGGCATCGCGGCAACGAG 2304  
 Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700  
 QY 2305 CAGATCGACAAGCTGTGAGCAAGGCGATCCGCAAGGTGCTGTCTGACCGCATCGAT 2364  
 Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720  
 RESULT 3  
 US-08-463-210-9  
 ; Sequence 9, Application US/08463210  
 ; Patent No. 6001977  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHANG, Nancy T.  
 ; APPLICANT: GALLO, Robert C.  
 ; APPLICANT: WONG-STAAAL, Flossie  
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morgan & Finnegan, L.L.P.  
 ; STREET: 345 Park Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10154-0053  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/463,210  
 ; FILING DATE: 05-JUN-1995  
 ; CLASSIFICATION: 436  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 06/693,866  
 ; FILING DATE: 23-JAN-1985  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 06/659,339  
 ; FILING DATE: 10-OCT-1984  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Serunian, Leslie A.  
 ; REGISTRATION NUMBER: 35,353  
 ; REFERENCE/DOCKET NUMBER: 2026-4193US2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 758-4800  
 ; TELEFAX: (212) 751-6849  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1015 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: HTLV-III  
 ; FEATURE:  
 ; NAME/KEY: Protein  
 ; LOCATION: 1..1015  
 ; OTHER INFORMATION: /note= "pol protein of HTLV-III"  
 ; US-08-463-210-9  
 Alignment Scores:  
 Pred. No.: 2,82e-207 Length: 1015  
 Score: 3475.00 Matches: 654  
 Percent Similarity: 93.72% Conservative: 33  
 Best Local Similarity: 89.22% Mismatches: 28  
 Query Match: 75.99% Indels: 18  
 DB: 3 Gaps: 4  
 US-09-610-313B-31 (1-2463) x US-08-463-210-9 (1-1015)

QY 220 TTCCTCCGAGGACCTGGCTTCCCTCCCGAGGCAAGGCCCGCGAGTTC----- 267  
Db 1 PhePheArgGluAaspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 268 -----CCGAGCGAGCAGAACCGCGCCACAGCCCGCCACCGCAGC 303  
Db 21 ThrArgAlaAenSerProThrIleSerSerGluGlnThrArgAlaAenSerProThrArg 40  
QY 304 CGCGAGCTGCAGGTG-----CGCGCGACACAACCCCGCAGCAGCGCGCGCGAGCGC 357  
Db 41 ArgGluLeuGlnValTTPGlyArgAaspAenAenSerProSerGluAlaGlyAlaAaspArg 60  
QY 358 CAGGCGACCCCTG-----AATCTCCCGCAGATCACCTGTGTGGCGCGCCCTGGTGAGC 411  
Db 61 GlnGlyThrValSerPheAenPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
QY 412 ATCAAGGTGGCGCGCCAGATCAAGGAGCGCTGTGTGGACACCGCGCGCGAGCACCGTG 471  
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAaspThrGlyAlaAaspThrVal 100  
QY 472 CTGAGGAGATGAGCTGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGC 531  
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120  
QY 532 TTCTCAAGGTGCGCCAGTACGACAGATCTGTATCGAGATCTCGCGCAAGAGGCCATC 591  
Db 121 PheIleLysValArgGlnTyrAaspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
QY 592 GGCACCGTGTGATCGCGCCCGCCCGTGAACATCATCGCGCGCCCAACATGCTACCCAG 651  
Db 141 GlyThrValLeuValGlyProThrProValAenIleIleGlyArgAenLeuLeuThrGln 160  
QY 652 CTGGCTGCACCTGAACTTCCCATCAGCCCCATCGAGACCGTGGCCCGTGAAGCTGAAG 711  
Db 161 IleGlyCysThrLeuAenPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CCGGCGATGAGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCC 771  
Db 181 ProGlyMetAepGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCGCCATCTCGAGGAGATGGAGAGAGGCGCAGATCACCAAGATCGGCCCGCAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCTTACAAACCCCGCTGTCGCCATCAAGAGAGAGCAGCAGCACAAGTGGCGCAAG 891  
Db 221 AsnProTyrAenThrProValPheAlaIleLysLysLysAaspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGC 951  
Db 241 LeuValAaspPheArgGluLeuAenLysArgThrGlnAaspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCCACC CGCGCTGCTGAAGAAGAGAGAGCTGACCGTGGAGTGGCGAC 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAaspValGlyAasp 280  
QY 1012 GCCTACTTCAGCTGCGCCCTGAGAGGACTTCGCGCAAGTACACGCTTCACCATCCCC 1071  
Db 281 AlaTyrPheSerValProLeuAaspGluAaspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACACGAGACCCCGCGCATCCGCTACAGTACACAGTGTGCTGCCCGCGGTGG 1131  
Db 301 SerIleAenAenGluThrProGlyIleArgTyrGlnTyrAenValLeuProGlnGlyTrp 320  
QY 1132 AAGGCGACCCCGCATCTCCAGAGCAGCATGACCAAGATCTCGGAGCCCTTCGCGCC 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCCGAGATCGGTATCTACAG-----GCCCGCTGTACGTGGGCGAGCAGCTG 1245  
Db 341 GlnAenProAaspIleValIleTyrGlnTyrMetAaspAaspLeuTyrValGlySerAaspLeu 360

QY 1246 GAGATCGCGCGAGCACCGCGCCCAAGATCGAGAGCTGCGCAAGACCACTCTCTGGCTGGGC 1305  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuAargTrpGly 380  
QY 1306 TTCAACACCCCGCAAGAAGCACCAGAAGAGAGCCCTTCTCTGTGTGATGGCTACGAG 1365  
Db 381 LeuThrThrProAaspLysLysHisGlnLysGluProPheLeuLeuTrpMetGlyTyrGlu 400  
QY 1366 CTGACACCCCAACAAGTGGACCGCTGAGCCCATCGAGTGGCCGAGAGAGAGAGTGGACC 1425  
Db 401 LeuHisProAaspLysTrpThrValGlnProIleValLeuProGluLysAaspSerTrpThr 420  
QY 1426 GTGAACACATCCAGAAGCTGGTGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGC 1485  
Db 421 ValAenAaspIleGlnLysLeuValGlyLysLeuAenTrpAlaSerGlnIleTyrProGly 440  
QY 1486 ATCAAGGTGGCGAGCTGTGCAAGCTGCGCGCGCGCAAGGCCCTGACCGACATCGTG 1545  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuAargGlyThrLysAlaLeuThrGluValIle 460  
QY 1546 CCCCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAACCCGAGATCTCTCGCGAGGCC 1605  
Db 461 ProLeuThrGluGluAlaGluLeuGluAlaGluAenArgGluIleLeuLysGluPro 480  
QY 1606 GTGACCGCGCTGTACTAGCACCCCGACAGACCTGTGTGGCGCAGATCCAGAGCAGGC 1665  
Db 481 ValHisGlyValTyrTyrAaspProSerLysAaspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1666 CACGACAGTGGACCTACCAAGATCTACAGAGCGCTTCAAGAACCTGAAGCCGCGAAG 1725  
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAenLeuLysThrGlyLys 520  
QY 1726 TACGCCAAGATGCGCCCGCCACCAACAGACGTGAAGCAGCTGACCGAGCGCGTGCAG 1785  
Db 521 TyrAlaAargMetArgGlyAlaHisThrAenAaspValLysGlnLeuThrGluAlaValGln 540  
QY 1786 AAGATCCCGCTGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATC 1845  
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGCGCCACTGATCCCGAG 1905  
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1906 TGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGTGACCTGACCTGGAGAGGAGGCC 1965  
Db 581 TrpGluPheValAenThrProProLeuValLysLeuTrpTyrGlnLeuLysGluPro 600  
QY 1966 ATCATCGCGCGAGACCTTCTAGCTGAGCGCGCGCCCAACCCGCGAGACCAAGATCGGC 2025  
Db 601 IleValGlyAlaGluThrPheTyrValAaspGlyAlaAlaAenArgGluThrLysLeuGly 620  
QY 2026 AAGSCGCGTACGTGACCGCGCGCGCGCAGAGATCGTGACCTGACCGAGACCAACC 2085  
Db 621 LysAlaGlyTyrValThrAenLysGlyArgGlnLysValValProLeuThrAenThrThr 640  
QY 2086 AACCAGAGACCCAGCTGCGAGCGCCATCCAGCTGCGCCCTGCGAGCAGCGCGAGGAGTG 2145  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAenSerGlyLeuGluVal 660  
QY 2146 AACATCGTGCAGACCGCAGTACGCCCTGGGCATCATCGAGCGCCCGCAGACCAAGAGC 2205  
Db 661 AsnIleValThrAaspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAaspLysSer 680  
QY 2206 GAGAGCGAGCTGTGAACACAGATCATCGACGCTGATCAAGAGGAGAGGCTGACTGCTG 2265  
Db 681 GluSerGluLeuValAenGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700  
QY 2266 AGCTGGTGTCCCGCCCAAGAGGCGCATCGCGCGCAACCGAGCAGATCGCAACTGCTGAGC 2325  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlnGlnValAaspLysLeuValSer 720  
QY 2326 AAGGCGCATCCGCAAGGTGCTGTTCTTCTGGACGGCATCGAT 2364



```
QY 1726 TACGCCAAGATGCGCACCGCCCAACACCAACGACGTGAAGAGCTGACCGAGCGCGTGCAG 1785
Db 521 TyrAlaargMetargGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1786 AAGATCGCCATGGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTCGGCTCCGCATC 1845
Db 541 LysileThrThrGluSerileVallelrrpGlyThrProLysPheLysLeuProlle 560
QY 1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGGCCACCTGATCCCGAG 1905
Db 561 GlnLysGluThrTrpGluThrTrpThrGluThrTrpGlnAlaThrTrpIleProGlu 580
QY 1906 TGGAGTTCGTGAACACACCCCGCTGTGAAGCTGGTACCGAGCTGGAGAGAGGCC 1965
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrrGlnLeuGluLysGluPro 600
QY 1966 ATCATCGCGCGGAGACCTTCTAGTGGACCGCGCGGCGGAGATCGTGAGCCTGACCGACACCC 2025
Db 601 IleValGlyAlaGluThrPheTyrrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2026 AAGCCCGGCTACGTGACCGACCGCGGCGGCGGAGATCGTGAGCCTGACCGACACCC 2085
Db 621 LysAlaGlyTyrrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThr 640
QY 2086 AACAGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCGAGGACCGCGGAGGAGTG 2145
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2146 AACATCGTGACCGACAGCAGTACGCTGGGCGATCATCCAGGCCCGCGGCAAGAGC 2205
Db 661 AsnIleValThrAspSerGlnTyrrAlaLeuGlyIlelleGlnAlaGlnProAspLys 680
QY 2206 GAGACGAGCTGGTGAACACGATCATCGAGCAGCTGATCAAGAAGAGAGAGGTGTACTG 2265
Db 681 GluSerGluLeuValAsnGlnIlelleGluGlnLeuIleLysLysGluLysValTyrrLeu 700
QY 2266 AGCTGGTGGCGGCCCAAGGGGATCGCGGCAACGAGGAGATCGACAAGCTGGTGCAGC 2325
Db 701 AlaTrpValProAlaHisLysGlylleGlyLysGlnGlnValAspLysLeuValSer 720
QY 2326 AAGGCGATCCGAGGTGCTGCTTCTGCGGCGATCGAT 2364
Db 721 AlaGlylleArgLysIleLeuPheLeuAspGlylleAsp 733
```

## RESULT 5

```
US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,028
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
```

```
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1015
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
; US-08-463-028-9
```

```
Alignment Scores:
Pred. No.: 2,826-207 Length: 1015
Score: 3475.00 Matches: 654
Percent Similarity: 93.72% Conservatives: 33
Best Local Similarity: 89.22% Mismatches: 28
Query Match: 75.99% Indels: 18
DB: 4 Gaps: 4

US-09-610-313B-31 (1-2463) x US-08-463-028-9 (1-1015)

QY 220 TTCTTCGCGAGGACCTGGGCTTCCCCAGGGCAAGGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 268 -----CCCAGCGAGAGAACCCGCGGCAACAGCCGCGCAACGCCACCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGGCGACAAACCCCGCGAGCGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGCGACCCCTG-----AACTTCCCCAGATCACCTGTGGCAGCGCCCTCGTGCAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAGGTGGCGCGCGAGATCAAGAGGCCCTGTGGACACCGCGCGCGCGAGCACCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTTCGCGCAACTGGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
QY 532 TTCATCAAGGTGCGCCAGTACGACACAGATCCTGTATCGAGATCTCGCGCAAGAAGCCCATC 591
Db 121 PheIleLysValArgGlnTyrrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGGCGCCCGCCCGTGAACATCATCGCGCGCGCGCATCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIlelleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCCCTGAACCTTCCCATCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
```

QY 712 CCGGCATGACGGCCCAAGGTGAAGCTGGCCCTGACCGAGGAGAAAGATCAAGGCC 771  
DB 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCGCCATCTCGAGGAGATGGAGAAGAGGCGCAAGATCAACAAGATCGGCCCGAG 831  
DB 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTACACACCCCGGTTCGCCATCAAGAAAGACAGACCAAGTGGCGCAAG 891  
DB 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTCGGAAGGTGCAGCTGGGC 951  
DB 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCACCCCGCCGCTGAAGAAGAAGAGCGTGACCGTGTGGAGCTGGCGGAC 1011  
DB 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTCAGCTGCCCTCGAGGAGACTTCGCGAAGTACACCGCTTCACCATCCCC 1071  
DB 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACAAGAGACCCCGGCTCCGCTACCAAGTACAAGCTGCTGCCCGCCAGGGCTGG 1131  
DB 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGGCAGCCCGACATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCGCGCC 1191  
DB 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCGAGATCGTGATACCAG-----GCCCGCTGTACGTGGGCGAGGACCTG 1245  
DB 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
QY 1246 GAGATCGGCAGCACCGCGCAAGATCGAGGAGCTGGCAAGCACCTGCTGCGCTGGGGC 1305  
DB 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380  
QY 1306 TTCACCAACCCCGCAAGAAGACCAAGAGAGCCCGCTTCCTGTGGATGGGCTACGAG 1365  
DB 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400  
QY 1366 CTGCACCCCAAGTGGACGTGGCGAGCCCATCGAGCTGCCGAGAGCGAGAGAGCTGGACC 1425  
DB 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1426 GTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGCGCCAGCAGATCTACCCCGGC 1485  
DB 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
QY 1486 ATCAAGTGGCCCGCATGTGCAAGCTGCTGCGGCGGCCAAGCGCTTCAAGACATCGTG 1545  
DB 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1546 CCCTGACCGAGAGGCGGAGCTGGAGCTGCCGAGAGCGCGAGATCCTGCGCGAGCCC 1605  
DB 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1606 GTGCACCGGCTGTACTACGACCCCGACAGGACCTGGTGGCCGAGATCCAGAAGAGAGGGC 1665  
DB 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1666 CACGACCAAGTGGACCTTACCAAGATCTACCAAGAGCCCTTCAAGAACCTGAAGACGGCAAG 1725  
DB 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1726 TAGCCCAAGATGCCACCGCCACACCAAGAGCTGAAGGAGCTGACCGAGGCGCTGCAG 1785  
DB 521 TyrAlaAspMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1786 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATC 1845

DB 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1846 CAGAAGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAG 1905  
DB 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1906 TGGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGGTACCAGCTGGAGAGAGGAGGCC 1965  
DB 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600  
QY 1966 ATCATCGGCGCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGC 2025  
DB 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2026 AAGCGCGGTACGTGACCGACCGGCGCGCGCGCAGAGATCGTGGCTGACCGAGACACC 2085  
DB 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2086 AACCAAGAAGACCGAGCTGCAGGCGCATCCAGCTGCGCTGCAGACACAGCGCAGAGGTG 2145  
DB 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2146 AACATCGTGACCGACAGCCAGTACCGCTGGGCATCATCCAGCGCCCGACCGCAAGAGC 2205  
DB 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2206 GAGAGGAGCTGTGTGAACCGATCATCGACGACCTGATCAAGAGAGAGGTGTACTGT 2265  
DB 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700  
QY 2266 AGCTGGTGTGCCCGCCCAAGGCGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325  
DB 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2326 AAGGCGATCCGCAAGTGTCTTCTTGGAGCGCATCGAT 2364  
DB 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 6  
US-07-743-357-2  
; Sequence 2, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900

```

; TELFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: BH102
US-07-743-357-2

Alignment Scores:
Pred. No.: 2,82e-207 Length: 1016
Score: 3475.00 Matches: 654
Percent Similarity: 93.72% Conservative: 33
Best Local Similarity: 89.22% Mismatches: 28
Query Match: 75.99% Indels: 18
DB: 2 Gaps: 4

US-09-610-313B-31 (1-2463) x US-07-743-357-2 (1-1016)

Qy 220 TTCTTCGCGAGGACCTGGCCCTTCCCGCAGGCAAGGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAsePLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
Qy 268 -----CCGAGCGAGCAGAACCGCGCCACAGCCCGCCACAGCCCGCCACAGC 303
Db 21 ThrArgAlaAenSerProThrIleSerSerGluGlnThrArgAlaAenSerProThrArg 40
Qy 304 CGCGAGCTGCAGGTG-----CGCGCGCACAACCCCGCAGCGAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAenAenSerProSerGluAlaGlyAlaAspArg 60
Qy 358 CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCGAGCCCGCCCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAenPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
Qy 412 ATCAAGTGGCGCGCCAGATCAAGAGCGCCCTGTGCACACCGCGCGCGCGAGCACCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAsePThrGlyAlaAsePThrVal 100
Qy 472 CTGAGGAGATGACCTCGCCCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluLeuMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
Qy 532 TTCATCAAGTGGCCAGTACGACCATCTGTGATCGAGATCTGCGCAAGAGCGCCATC 591
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
Qy 592 GGCACCGTGTGATCGCGCGCCACCGCGTGAACATCATCGCGCGCCACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAenIleIleGlyArgAenLeuLeuThrGln 160
Qy 652 CTGGGCTGACCCCTGAATCTCCCATCAGCCCATCAGACCGCGCGCGCGTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAenPheProIleSerProIleGluThrValProValLysLeuLys 180
Qy 712 CCCGGCATGAGCGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
Qy 772 CTGACCCCATCTCGAGGAGATGGAGAGGCGCAAGATCACCAAGATCGGCCCGCCAG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
Qy 832 AACCCCTTACAAACCCCGCTGTTCGCCATCAAGAGAGAGACACACCAAGTGGCGCAAG 891
Db 221 AsnProIlyrAenThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

```

```

Qy 892 CTGCTGACTTCCGCGAGCTGAAACAAGGCGACCCAGGACTTCTGGAGGTCGAGCTGGGC 951
Db 241 LeuValAspPheArgGluLeuAenLysArgThrGlnAsePheTrpGluValGlnLeuGly 260
Qy 952 ATCCCCACCCCGCGCGCTGAAAGAAGAGCGTGCAGCGTGTGCGAGCTGGCGCAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAsePValGlyAsp 280
Qy 1012 GCCTACTTTCAGCGTCCCTTCCGAGGACTTCCGCAAGTACACCGCTTCCACATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAsePLeuAsePheArgLysTyrThrAlaPheThrIlePro 300
Qy 1072 AGCATCAACAGAGACCCCGCGCATCGCTACAGTACAGCTGCTGCCCGCGCGCTGG 1131
Db 301 SerIleAenAenGluThrProGlyIleArgTyrGlnTyrAenValLeuProGlnGlyTrp 320
Qy 1132 AAGGCGAGCCCGCAGCATCTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
Qy 1192 CGCAACCCCGAGATCGTGATCTTACCAG-----GCCCGCTGTGTGGCGAGCGACCTG 1245
Db 341 GlnAenProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
Qy 1246 GAGATCGCGCAGACCCCGCGCAGATCGAGAGCTGGCGAAGACCTGCTCGCTGGCGC 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380
Qy 1306 TTCAACACCCCGCAAGAAGCAGACAGAGAGCCCGCTTCTGTGTGATGGGTACTACGAG 1365
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400
Qy 1366 CTGCACCCCGCAGAGTGGAGCCGTGCAGCCCATCGAGTGCCTCCGAGAGAGAGCTGGACC 1425
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAsePTrpThr 420
Qy 1426 GTGAACAGATCCAGAGCTGTGGGCAAGCTGAACCTGGCGCGCAGCAGATCTACCCGCGC 1485
Db 421 ValAenAspIleGlnLysLeuValGlyLysLeuAsePTrpAlaSerGlnIleTyrProGly 440
Qy 1486 ATCAAGTGGCGCAGCTGTCAAGCTGTGCGCGCGCCAGGCGCTGACCGACATCGTG 1545
Db 441 IleLysValArgGlnLeuLysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
Qy 1546 CCCTTGACCGAGGAGCGCGAGCTGGAGCTGCGCGAGAACCGCGAGATCTCTCGCGAGCCC 1605
Db 461 ProLeuThrGluGluAlaGluLeuLeuAlaGluAsePTrpGluIleLeuLysGluPro 480
Qy 1606 GTGCACCGCTGTACTAGACCCCGCAGACCTGTGTGGCGCGAGATCCAGACAGCGC 1665
Db 481 ValHisGlyValTyrTyrAspProSerLysAsePLeuIleAlaGluIleGlnLysGlnGly 500
Qy 1666 CACGACGAGTGGACCTACAGATCTACAGAGCCCTTCAAGACCTGAAGACCGCGCAAG 1725
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnLysProPheLysAenLeuLysThrGlyLys 520
Qy 1726 TAGCGCAAGATGGCGACCCCGCCACCAACAGCAGCTGAAGCAGAGCTGACCGCGCGCTGAG 1785
Db 521 TyrAlaArgMetArgGlyAlaHisThrAenAsePValLysGlnLeuThrGluAlaValGln 540
Qy 1786 AAGATCGCCATGGAGAGATCGTGATCTGGGGCAAGACCCCGCAAGTTCGCGCTGCCCATC 1845
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
Qy 1846 CAGAGGAGACCTGGGAGACCTGTGACCGCTACTCTGGCGAGCGCCACCTGATCCCCGAG 1905
Db 561 GlnLysGluThrTrpGluThrTrpThrGlyThrTrpGlnAlaThrTrpIleProGlu 580
Qy 1906 TGGAGTTCGTGAACACCCCGCCCTGCTGAAGCTGTGGTACCACTGCGAGAGAGAGCGCC 1965
Db 581 TrpGluPheValAenThrProProLeuValLysLeuTrpTrpGlnLeuGluLysGluPro 600
Qy 1966 ATCATCGGCGCGAGACCTTCTTACGTGGAACCGCGCGCCCAACCGCGAGACCAAGATCGGC 2025

```



```
Db 601 ILeValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2026 AAGCGCGGTACTGACCGACCGGGCGGCGAGAGATCGTAGCTGACCGACGACACC 2085
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValProLeuThrAsnThrThr 640
QY 2086 AACGACAGACCGAGCTGCAGGCGCATCCAGCTGCGCCAGGACAGCGCAGCGAGGTG 2145
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2146 AACATCGTGACCGACAGCCAGTAGTACGCCCTGGGCATCATCCAGCGCCAGCCCGCAAGAGC 2205
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2206 GAGAGCGAGCTGTGTAACAGATCATCGACGACGCTGATCAGAGAGAGGTGTACCTG 2265
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2266 AGCTGGGTGCCCGCCACAGGCGCATCCGCGGCAACAGGACAGATCGACAAGCTGTGAGC 2325
Db 701 AlaTrpValProAlaHisLysGlyIleGlyLysGlnGluGlnValAspLysLeuValSer 720
QY 2326 AAGCGCATCCGCAAGTGCTGTCTCTCGACGCGCATCGAT 2364
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7
US-09-309-572-17
; Sequence 17, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCES: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-309-572-17

Alignment Scores:
Pred. No.: 4,32e-207 Length: 1003
Score: 3472.00 Matches: 652
Percent Similarity: 95.15% Conservative: 34
Best Local Similarity: 90.43% Mismatches: 29
Query Match: 75.92% Indels: 6
DB: 4 Gaps: 3

US-09-610-313B-31 (1-2463) x US-09-309-572-17 (1-1003)
QY 220 TTCTTCCGCGGAGGACCTGGCGCTTCCCGGAGGCAAGCGCCGCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCAAACGCCACCGCCAGCGCGAGCTGAGGTG-----CGCGCGCAACACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCCAGCGCACCTCG-----AACTTCCCGCCAGATCACCC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCGCTGTGTGAGCATCAAGTGGCGCGCGCGATCAAGGAGCGCCCTGCTG 447
```

```
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGGCGCGACGACACCGCTGCTGGAGGAGATGAGCTGCCCGGCAAGTGAAGACCCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGGCGGATCGCGGCTTTCATCAAGGTGCGCGAGTACGACGACATCTCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTGGCGCAAGAGGCCATCGCACCGCTGCTGATCGGCGCCACCCCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATCGTACCCAGCTGGGCTGCACCTGAACCTTCCCATCAGCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTCCCGTGAAGCTGAAGCCCGCATCGACGCGCCCAAGGTGAAGCAAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGAGAGATCAAGGCCCTGACCGCATCTGCGAGGAGATGAGAGAGAGGCGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCCCGAGAACCCCTACAACACCCCGCTGTCGCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCCCGAGCTGAACAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGAGGTGACGTGGGCATCCCGCCACCGCGCGCTGAGAGAGAGAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260
QY 988 GTGACCGTGTGGACGTGGCGCGACGCTACTTTCAGCGTCCCGTGGACGAGGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCTTACCATCCCGATCAACACGAGACCCCGCGCATCCGCTACCAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGCTGCCAGGCTGGAGGCGCGCCCGAGCATCTTCCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320
QY 1168 AAGATCTGGAGCGCTTCCGCGCGCCGCAACCCCGAGATCGTGTATCCAG-----GCC 1221
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY 1222 CCCTGTAGCTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCCCAAGATCGAGGAGCTG 1281
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 CGCAACGACCTGCTGGCGCTGGGCTTCAACCCCGCGCAAGAGCAGCAGAGAGAGAGCGCC 1341
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCCTTCTGTGATGGCTACGAGCTGCACCCGCAAGTGGACCGCTGACCGCCATCGAG 1401
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1402 CTGCCCGAGAGAGAGCTGGACCGTGAACGACATCAGAAAGCTGTGTGGCGCAAGCTGAAC 1461
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGCGCAGCGCATCTACCCCGCGCATCAAGTGGCGCGCGCTGCAAGCTGTGCGCGGC 1521
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
```

```
QY 1522 GCCAAGGCCCTGACGACATCGTGGCCCTGACCCGAGAGGCGCGAGCTGGAGCTGGCCGAG 1581
Db 441 ThrLysAlaLeuThrGluValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
QY 1582 AACCGCGAGATCTCGCGAGCGCGTGTACCGCGCTGTACTACGACCCCGAGCAAGGACTG 1641
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1642 GTGGCCGAGATCCAGAAGCAGGCGCACGACAGTGGACCTTACCAGATCTACCAGAGGCC 1701
Db 481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
QY 1702 TTCAGAAGACCTGAAGACCGGCAAGTACGCCAAGATGGCCACCGCCACACCAAGCAGCTG 1761
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520
QY 1762 AAGCAGCTGACCGAGGCGGTGACAGAATGCCATGCCATGGAGACATCGTATCTGGGGCAAG 1821
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTyrGlyLys 540
QY 1822 ACCCCCAAGTTCGCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACTAC 1881
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTyrGluAlaTyrTyrThrGluTyr 560
QY 1882 TGGCAGGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1941
Db 561 TrpGlnAlaThrTyrIleProGluTyrPheValAlaThrGluSerIleValIleTyrGlyLys 580
QY 1942 TGGTACAGCTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTAGTGGAGCGGCC 2001
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGCGGCGCGCAGAG 2061
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
QY 2062 ATCGTGACCTGACCGAGACCAACACGAGACCGAGCTGCAGGCGCATCCAGCTGGCC 2121
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2122 CTGCAGACAGCGCGAGGAGTGAACATCGTGGCCGAGACCTGACCGACCGCTGGGCATC 2181
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2182 ATCCAGGCCCGCAGCCGACAGGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTG 2241
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAAGGAGAAGGTGTACCTGAGCTGGGTGCGCCGCCCAAGGGCATCGCGGCAAC 2301
Db 681 IleLysLysGluLysValTyrLeuAlaTyrValProAlaHisLysGlyIleGlyAsn 700
QY 2302 GAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCCTGGAGCGGCATC 2361
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
Db 721 Asp 721
```

## RESULT 8

```
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718.096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
```

```
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-718-096-17
```

## Alignment Scores:

Pred. No.:	4.32e-207	Length:	1003
Score:	3472.00	Matches:	652
Percent Similarity:	95.15%	Conservative:	34
Best Local Similarity:	90.43%	Mismatches:	29
Query Match:	75.92%	Indels:	6
DB:	4	Gaps:	3

US-09-610-313B-31 (1-2463) x US-09-718-096-17 (1-1003)

```
QY 220 TTCTTCCGCGAGGACCTGGCTTCCCCAGGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCCAACGCCCCACCGCGCGAGCTGCAGGTG-----CGCGGCGCAACACCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTyrGlyArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCGCGAGCGCACCTG-----AACTTCCCCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGGCGCCCTGGTGTGAGCATCAAGTGTGGCGGCGCAGATCAAGAGGCGCTGTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCGAGCACCGCTGCGAGGAGATGAGCTGCCCGCGCACTGGAGGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGGCGCCAGTACGACCATCTCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeu 120
QY 568 GAGATCTGGCGCAAGAGGCGCATCGGCACCGTGTGTGATCGGCCCGCCACCGCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGTGTGACCCAGCTGGGTGACACCTGCACTTCCCATCAGCCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCTGCGCGTGAAGCTGAAGCGCGCATGACCGCGCGCGCCCAAGGTGAAGCAGTGGCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTyrPro 180
QY 748 CTGACCGAGGAGAGATCAAGAGCGCCCTGACCGCCCATCTGCGGAGGAGATGGAGAGAGGC 807
Db 181 LeuThrGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGAGAACCTTACACACACCCCGTGTGGCCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGACTTCCCGAGCTTCCCGAGCTGAACAAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGAGGTGTCAGCTGGGCATCCCCCAACCCCGCGCGCTTGAAGAAGAAGAGAGC 987
```

```
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysSer 260
QY 988 GTGACCGTGTGACGCGGCGAGCGCTACTTACGGTGGCCCTGACGAGGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCTTACCATCCCGACGATCAACCAAGCAGACCCCGGCATCCGCTACCAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAAGCTGTGCTGCCCGAGGCTGGAAGGCGAGCCCGACGATCTTCCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320
QY 1168 AAGATCTGTGAGCCTTCCGCGCGCGCAACCCGAGATCGTGATCTACCG-----GCC 1221
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY 1222 CCCTGTACGTGGCGCAGCACTGGAGATCGGCAGCACCGCCCGCAAGATCGAGGAGCTG 1281
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 CGCAAGCACCTGTGCGCTGGCGCTTCCACCACCCCGCACAAAGAACCCAGAGAGAGGCC 1341
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCTTCTGTGGATGGCTACGAGTGCACCCGACCAAGTGGACCGTGCGAGCCCATCGAG 1401
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1402 CTGCCCGAGAGGAGCTGGACCGTGAACAGCATCCAGAGCTGTGGCGCAAGCTGGAAC 1461
Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGGCCAGCGAGATCTACCCCGCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGC 1521
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1522 GCCAAGCCTGACCGACATCGTGCCTCGCTACCGAGAGCGCGAGCTGGAGCTGGCGGAG 1581
Db 441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuValAlaGlu 460
QY 1582 AACCGGAGATCTGCGCGAGCGCGTGCACGGCTGTACTACGACCCCGACAGGAGCCTG 1641
Db 461 AsnArgGluIleLysGluProValHisGlyValTyrTrpAspProSerLysAspLeu 480
QY 1642 GTGCGCGAGATCCAGAGCAGGCGCCACGACAGTGGACCTTACGAGATCTACAGGAGCCC 1701
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
QY 1702 TTCAAGACCTGAAGCCGCAAGTAGCCCAAGATGGCACCGCCGACCAACAGCAGCTG 1761
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520
QY 1762 AACGAGCTGACCGAGCGCGTGCAGAGATCGCATCGAGAGCATCGTATCTGGGCGCAAG 1821
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540
QY 1822 ACCCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTTGGGAGACCTTGGTGGACCGACTAC 1881
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTyr 560
QY 1882 TGGCAGGCCACTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGTGGTGAAGCTG 1941
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1942 TGGTACGAGTGGAGAGGAGCCATCATCGGCGCGAGACCTTCTACGTGGAGCGCGCC 2001
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCGCGAGACCAAGATCGGACGCGCGGTACTGTGACCGACCGGCGCGCGAGAAG 2061
```

```
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
QY 2062 ATCGTGAGCTGCACCGAGACCCAGACCCAGACAGAGCTGCAGGCCATCCAGCTGGCC 2121
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2122 CTGCAGGACAGCGGCGACGAGGTGAACATCGTGACCGACGAGCCATCCCTGGCGCATC 2181
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2182 ATCCAGGCCCGCCGACCAAGAGCGAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTG 2241
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAGGAGAGGTGTACCTGAGCTGGTGGCGCCGACCAAGGGCATCGCGCGCAAC 2301
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2302 GAGCATCGACCAAGCTGTGTGAGCAAGGCGCATCCGCAAGGTGTCTTCTGAGCGGCATC 2361
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
Db 721 Asp 721
RESULT 9
US-07-743-357-5
; Sequence 5, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: BRU
```

US-07-743-357-5

## Alignment Scores:

Pred. No.: 6.64e-207 Length: 1016  
Score: 3469.00 Matches: 652  
Percent Similarity: 93.86% Conservative: 36  
Best Local Similarity: 88.95% Mismatches: 27  
Query Match: 75.86% Indels: 18  
DB: 2 Gaps: 4

US-09-610-313B-31 (1-2463) x US-07-743-357-5 (1-1016)

QY 220 TTCTTCCGAGGACCTGGCTTCCCGAGGCGAGCCCGGAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyAlaArgGluPheSerSerGluGln 20  
QY 268 -----CCGAGCGAGCAGAACCCGCGCAACAGACCCCGACAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
QY 304 CGCGAGTGCAGGTG-----CGCGGGGACAAACCCCGCAGCGAGCGCGCGCGAGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArg 60  
QY 358 CAGGGCACCCCTG-----AAGTTCCCGAGATCACCCTGTGGCAGCGCCCTGTGTGAC 411  
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
QY 412 ATCAAGGTGGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGCAGCACACCGTG 471  
Db 81 IleIysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100  
QY 472 CTGGAGGAGATGAGCTCTCCCGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGGC 531  
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyGly 120  
QY 532 TTGATCAAGTGGCGCGCAGTAGCAGATCCTGATCGAGATCTCGCGGCAAGAGCCATC 591  
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
QY 592 GGCAACGTGTGATCGCGCGCCCGCAGCATCATCGCGCGCAACATGCTGACCCAG 651  
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
QY 652 CTGGGCTGCACCTGAACTTCCCATCAGCCCGCATCGAGACCGTCCCGTGAAGCTGAAG 711  
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CCGCGCATGGACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCC 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysAla 200  
QY 772 CTGACCGCATCTCGGAGGAGATCGAAGAGGCGCAAGATCACCAAGATCGGCGCCGAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTACAAACCCCGCTTCCCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAG 891  
Db 221 AsnProIleThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCCCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGC 951  
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCGCACCCCGCGCTCAAGAAGAAGAGCGTGAACCGTGGAGCTGGCGCGGAC 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTCCGCTGCGCTGGAGGAGACTTCCGCAAGTACACCGCTTCCACCATCCCC 1071  
Db 281 AlaIlePheSerValProLeuAspGluAspPheArgLysThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACACGAGACCCCGCGCATCGGTACAGTACAGTGTGCGCCCGCGAGGTGG 1131

Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGCGAGCCCGCATCTTCCAGAGCAGCATGACCAAGATCTCCGAGCCCTTCCGCGCC 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340  
QY 1192 CGCAACCCCGAGATCGTGATCTACAG-----GCCCGCTGTAGTGGGAGGAGACCTG 1245  
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
QY 1246 GAGATCGCGCAGCAGCCCGCAAGATCGAGAGCTGGCAAGACCTGCTGCTGGCGGC 1305  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380  
QY 1306 TTCAACCCCGCAGCAAGAGCAGAGAGGAGCCCGCTTCTGTGTGATGGGTACGAG 1365  
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400  
QY 1366 CTGACCCCGCAAGTGGACCGTGCAGCCCATCAGCTGCCGAGAGGAGAGTGGAGC 1425  
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1426 GTGAACGACATCCAGAGCTGTGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGC 1485  
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
QY 1486 ATCAAGGTGGCGCAGCTGTGCAAGCTGTGGCGCGCGCAGAGGCGCTGACGAGATCGTG 1545  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1546 CCGCTGACCGAGGAGCGAGCTGGAGTGGCCGAGAACCGCGAGATCTCTCGCGAGGCC 1605  
Db 461 ProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1606 GTGCACGCGGTGTACTAGCCCCAGCAGGAGCTGTGGCGCAGATCCCAAGCAGGCGC 1665  
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1666 CACGACGAGTGGACCTTACAGATCTACGAGAGCGCTTCAAGAACCTTGAAGACCGGCAAG 1725  
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnLysProPheLysAsnLeuLysThrGlyLys 520  
QY 1726 TAGCCCAAGATGGCCACCGCCCAACCAAGCTGAAGAGTGAAGCGGCGGTGAG 1785  
Db 521 TyrAlaArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1786 AAGATCCCATGGAGAGCATCTGATCTGGGCGAGACCCCAAGTTCGGCTGCCATC 1845  
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1846 CAGAAGGAGACCTGGGAGACCTGTGACCGACTACTGGGAGGCGCACCTGATCCCGAG 1905  
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1906 TGGAGTTCTGAACACCCCGCTGTGAGCTGTGGTACAGCTGGAGAGAGAGGCC 1965  
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600  
QY 1966 ATCATCGCGCGAGACTTCTAGTGGCGCGCGCCCAACCGCGAGACCAAGATCGGC 2025  
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620  
QY 2026 AAGCGCGCTACGTGACCGCGCGCGCGAGAGAGCTGTGAGCTGACCGAGCAGACCAAC 2085  
Db 621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValThrLeuThrAspThrThr 640  
QY 2086 AACCAAGAGCCGAGCTGCAGGCCATCCAGCTGCCCTGCGAGGACAGCGGAGAGGTG 2145  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2146 AACATCTGACCGACCGCAGTACCGCTTGGGATCATCCAGGCGCGCGCGCAGCAGAGC 2205

Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleGlnAlaGlnProAspLysSer 680  
QY 2206 GAGAGCGAGCTGTGACCATCATCGACAGCTGATCAAGAGGAGGTGACCTG 2265  
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGluLysValTyrLeu 700  
QY 2266 AGCTGGGTGCGCCCAACAGGGCATCGCGGCAACGAGCAGATCGACAGCTGTGAGC 2325  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2326 AAGGCGATCCGCAAGGTGCTTCTCGACGGCATCGAT 2364  
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733  
RESULT 10  
US-07-743-357-3  
; Sequence 3, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1016 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: BH5  
; US-07-743-357-3  
Alignment Scores:  
Pred. No.: 7.66e-207 Length: 1016  
Score: 3468.00 Matches: 652  
Percent Similarity: 93.72% Conservative: 35  
Best Local Similarity: 88.95% Mismatches: 28  
Query Match: 75.84% Indels: 18  
DB: 2 Gaps: 4  
US-09-610-313B-31 (1-2463) x US-07-743-357-3 (1-1016)

QY 220 TTCTTCCGCGAGGACCTGCCTTCCCCAGGCAAGCGCCCGAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 268 -----CCACGAGCAGAACCGCGCCCAACAGCCCCCACCAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
QY 304 CCGAGCTCCAGGTG-----CGCGCGACAACCCCGCAGCGAGCGCGCGAGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
QY 358 CAGGCGACCTCG-----AACTTCCCGCAGATCACCTCTGGCAGCGCCCTGTGTGAGC 411  
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
QY 412 ATCAAGTGGCGCGCGCAGATCAAGAGGCGCTCTGTGGACCGCGCGCCGACGACCGGTG 471  
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
QY 472 CTGGAGGATGAGCTGCGCGCGCGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGCGC 531  
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120  
QY 532 TTCATCAAGGTGCGCGCGCATGACGACCATCTGATCGAGATCTCGCGCAAGAGGCCATC 591  
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
QY 592 GGCACCGTCTGATCG 651  
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
QY 652 CTGGCGTGCACCTCGAATCTCCCATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711  
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CCGCGATGGAAGCG 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCGCATCTCGAGGAGATGGAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGluLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTACAACACCCCGTGTTCGCGCATCAAGAAGAAGAGAGGAGGAGGAGGAGGAGG 891  
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysLysLysLysLysLysLys 240  
QY 892 CTGGTGGACTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGCG 951  
Db 241 LeuValAspPheArgGluLeuAsnArgArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCGCACCCCG 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysLysLysLysLysLysLysLysLysLys 280  
QY 1012 GCCTACTTTCAGCTGCG 1071  
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACAAACGAGACCCCGCGCATCCGCTACAGTACAAACGTGTCGCCCGCGCGCTGG 1131  
Db 301 SerIleAsnAsnGluThrProGlySerGlyTyrGlnTyrAsnValLeuProGlnGlyTyr 320  
QY 1132 AAGGCGAGCG 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340  
QY 1192 CGCAACCCCGAGATCGTGTCTACCCAG-----GCCCGCTGTACGTGGCGCGCGCGCTG 1245  
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeu 360  
QY 1246 GAGATCGGCGCAGCACCG 1305





358 CAGGCGACCTG-----AACTTCCCCAGATCACCTGTGGCAGCGCCCTCGTGAGC 411  
|||||...  
61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
|||||...  
412 ATCAAGGTGGCGCGCCAGATCAAGGAGGCCCTCTCTGGACACCGCGCCCGACGACACCGTG 471  
|||||...  
81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
|||||...  
472 CTGGAGGAGATGACCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGC 531  
|||||...  
101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120  
|||||...  
532 TTCAATCAAGGTGGCGCGCCAGATCAAGACCATCTGATCGAGATCTGCGCGCAAGAGGCCATC 591  
|||||...  
121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
|||||...  
592 GGCACCGTGTGATCGCGCCCAACCCCGGTGAACATCATCGCGCGCAACATCTGTGACCCAG 651  
|||||...  
141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
|||||...  
652 CTGGGCTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTGGCCCGTGAAGCTGAAG 711  
|||||...  
161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
|||||...  
712 CCGGCATGAGCGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAATCAAGGCC 771  
|||||...  
181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
|||||...  
772 CTGACCGCCATCTGCGAGGAGATGGAGAAGGCGCAAGATCACCAAGATCGGCCCGCAG 831  
|||||...  
201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
|||||...  
832 AACCCCTAGAACACCCCGGTGTGGCCATCAAGAAGAGACAGCACCAAGTGGCGCAAG 891  
|||||...  
221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
|||||...  
892 CTGTGTGACTTCGCGAGCTGAACAGCGCACCGAGCTTCGCGAGGTGCACTGGGC 951  
|||||...  
241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
|||||...  
952 ATCCCCACCCCGCGCTGAAGAAGAAGAGCGTGACCGTGTGGAGCTGGGCGAC 1011  
|||||...  
261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
|||||...  
1012 GCCTACTTACGTGCCCCCTGGAGGAGACTTCGCGAAGTACACCGCTTCACCATCCCC 1071  
|||||...  
281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
|||||...  
1072 AGCATCAACAAACGAGACCCCGCGCATCCGCTACAGTACAAACGTGTGCTGCCCGAGGGCTGG 1131  
|||||...  
301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
|||||...  
1132 AAGGCGACCCCGACATCTTCCAGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGGCC 1191  
|||||...  
321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340  
|||||...  
1192 CGCAACCCCGAGATCGTGAATCTACCAG-----GCCCGCTGTGCTGGCGAGGACCTG 1245  
|||||...  
341 GlnAsnProAspIleValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeu 360  
|||||...  
1246 GAGATCGGCGAGCAGCGCGCAAGATCGAGAGCTGGCGAAGCACCTGCTGCGCTGGGCG 1305  
|||||...  
361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuAArgTrpGly 380  
|||||...  
1306 TTCACCAACCCCGCAAGAAGCACCAAGAGAGCCCGCTTCCTGTGGATGGGCTACGAG 1365  
|||||...  
381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400  
|||||...  
1366 CTGCACCCCGCAAGTGGACCGGTGCAGCCCATCGAGTGGCGAGGAGAGAGCTGGACC 1425  
|||||...  
401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
|||||...  
1426 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACCTGGGCGCACGATCTACCCCGGC 1485

Db ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
QY ATCAAGGTGGCGCGCCAGCTGTGCAAGCTGTGCGCGGCGCCAAAGCCCTGACGACATCGTG 1545  
Db IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluAlaIle 460  
QY CCCTCAGCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCGAGGCC 1605  
Db ProLeuThrGluGluAlaGluLeuLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY GTGCACGGCTGTACTACGACCCCAAGACCTGTGTGGCCGAGATCTCAGAAAGCAGAGGCC 1665  
Db ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY CAGCACCAGTGGACCTTACCAGATCTTACAGGAGCCCTTCAAGAACCTGAGACCGCGCAAG 1725  
Db GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY TAGCCCAAGATGCGCACCGCCCAACCAACGACGTCGAGCAGCTGACCGAGGCGCGTGCGAG 1785  
Db TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY AAGATCGCCATGGAGAGATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATC 1845  
Db LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY CAGAAGGACACTGGGAGACCTGTGGACCGACTACTGCGCGCCACCTGGATCCCCGAG 1905  
Db GlnLysGluThrTrpGluThrTrpThrGlyLysTrpGlnAlaThrTrpIleProGlu 580  
QY TGGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGTACCGCTGGAGAGAGAGGCC 1965  
Db TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600  
QY ATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCGCAACCGCGAGACCAAGATCGGC 2025  
Db IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGly 620  
QY AAGGCGGCTACTGTGACCGACCGCGCGCGCAGAGATCGTGTAGCTGACCGAGACCC 2085  
Db LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY AACCAAGAAGCCAGCTGCAGGCCATCCAGCTGGCCCTGCGAGCACGCGCGACGAGGTG 2145  
Db AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY AACATCGTCAACGACAGCCAGTACGCCCTGGGGCATCATCCAGGCCCGCCGACAGAGC 2205  
Db AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSer 680  
QY GAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTG 2265  
Db GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGlnLysValTyrLeu 700  
QY AGCTGGGTGCGCCCGCCCAAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGGTGAGC 2325  
Db AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY AAGGGCATCGCAAGTGTCTGTTCCTGGACCGCATCGAT 2364  
Db AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

## RESULT 12

US-07-743-357-10  
; Sequence 10, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: KIRBY EADES GALE BAKER  
 STREET: Box 3432, Station D  
 CITY: Ottawa  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: K1M 1H8  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/743,357  
 FILING DATE: 21-AUG-1991  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/CA90/00062  
 FILING DATE: 23-FEB-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gale, Edwin J.  
 REGISTRATION NUMBER: 28,584  
 REFERENCE/DOCKET NUMBER: 30924-2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (613) 237-6900  
 TELEFAX: (613) 237-0045  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1003 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 HYPOTHEetical: NO  
 FRAGMENT TYPE: internal  
 ORIGINAL SOURCE:  
 ORGANISM: Human immunodeficiency virus type 1  
 STRAIN: ELI  
 US-07-743-357-10

Alignment Scores:  
 Pred. No.: 2,56e-206 Length: 1003  
 Score: 3459,50 Matches: 646  
 Percent Similarity: 95.28% Conservative: 40  
 Best Local Similarity: 89.72% Mismatches: 29  
 Query Match: 75.65% Indels: 5  
 DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x US-07-743-357-10 (1-1003)

Qy	571	ATCTGGCGCAAGAGCGCCATCGGCACCTGCTGTGATCGGCGCCCGCCACCGCCGCGTGAACATCATC	630
Db	121	IleCysGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle	140
Qy	631	GGCGCAACATGCTGACCCAGCTGGCTGACCCCTGAACCTTCCCATCAGCCCATCGAG	690
Db	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
Qy	691	ACCGTGGCCGTGAAGCTGAAGCCCGGATGACCGCCCAAGGTGAAGCAGTGGCCCTG	750
Db	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
Qy	751	ACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAGATGGAAGAGGCGCAAG	810
Db	181	ThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys	200
Qy	811	ATCACCAAGATCGGCCCGGAGAACCTTACAAACCCCGCTGTCGCGCATCAAGAAGAAG	870
Db	201	IleSerArgIleGlyProGluAsnProTrpAsnThrProIlePheAlaIleLysLysLys	220
Qy	871	GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGAC	930
Db	221	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	240
Qy	931	TTCTGGGAGGTGACGCTGGGCATCCCCACCCCGCCGCTGAAGAGAAAGAGCGGTG	990
Db	241	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	260
Qy	991	ACCGTGTGGACGTGGCGGCGCTTCTTACGCGTCCCTGGAGGAGCTTCCGCAAG	1050
Db	261	ThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArgLys	280
Qy	1051	TACACCGCCTTCCACCATCCCGACATCAACAACAGAGACCCCGCGCATCCGCTACCACTAC	1110
Db	281	TyrThrAlaPheThrIleSerSerIleAsnAsnGluThrProGlyIleArgTrpGlnTyr	300
Qy	1111	AACGTGTGCCCCAGGGCTGGAGGGAGCGCCCGAGCATCTTCCAGAGCAGCATGACCAAG	1170
Db	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
Qy	1171	ATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGTACTACACAG	1224
Db	321	IleLeuGluProPheArgLysGlnAsnProGluMetValIleTrpGlnTyrMetAspAsp	340
Qy	1225	CTGTACGTGGCGACGACCTGGAGATCGGCAGCACCGCGCGCAAGATCGAGAGCTGCGC	1284
Db	341	LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg	360
Qy	1285	AAGCACCTGTGCTGGCGCTTCCACACCCCGACAGAACAGACCAAGAGGAGCCCGCC	1344
Db	361	GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro	380
Qy	1345	TTCTGTGGATGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTG	1404
Db	381	PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu	400
Qy	1405	CCCGAGAGGAGAGCTGGACCGTGACATCCAGAACGTGGTGGGCAAGCTGAACCTGG	1464
Db	401	ProGluLysGluSerTrpThrValAsnAspIleGlnAsnLeuValGluArgLeuAsnTrp	420
Qy	1465	GCACGACGATCTACCCCGGATCAGGTGGCGCAGCTGTGCAAGCTGTGCGGGCGGCC	1524
Db	421	AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr	440
Qy	1525	AAGCCCTGACCGACATCGTGGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGAGAAC	1584
Db	441	LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGluAsn	460
Qy	1585	CGGAGATCTTGGCGAGCCGCTGACGGGTGTACTACGACCCCGACGAGGAGCTGGTGG	1644
Db	461	ArgGluIleLeuLysGluProValHisGlyValTyrTrpAspProSerLysAspLeuIle	480
Qy	1645	GCCGAGATCCAGAGAGGCGCCACGACCACTGGACCTTACGAGTCTTACGAGGAGCCCTTC	1704

```
|||||
481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrglnIleTyrglnGluProPhe 500
QY 1705 AAGAAGCTGAAGACCGGAGTACGACGATGGACCGCCACACCAACGACGTGAAG 1764
Db 501 LysAsnLeuLysThrGlyLysTyraAlaArgMetArgGlyAlaHisThrAsnAspValLys 520
QY 1765 CAGCTGACGAGCGCGTGCAGAGATCGCCATCGATCGAGAGCATCTGTGGGGCAAGACC 1824
Db 521 GlnLeuAlaGluAlaValGlnArgIleSerThrGluSerIleValIleTrpGlyArgThr 540
QY 1825 CCCAAGTTCGCGCTGCCATCCAGAAAGAGACCTGGGAGACCTGGTGGACCCGACTACTGG 1884
Db 541 ProLysPheArgLeuProLysGlnLysGluThrTrpGluThrTrpAlaGluTyTrp 560
QY 1885 CAGGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCCCTGGTGAAGCTGTGG 1944
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580
QY 1945 TACGAGCTGGAGAGGAGCCATCATCGCGCGGAGACCTTCTACGTGGACGGCGCGCC 2004
Db 581 TyrGlnLeuGlnLysGluProIleIleGlyAlaGluThrPheTyValAspGlyAlaAla 600
QY 2005 AACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGCGCGCGGCAAGATC 2064
Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyraValThrAspArgGlyArgGlnLysVal 620
QY 2065 GTGAGCTGACCGAGACCAACACCAAGAGACCGAGCTGCAGGCCATCCAGTGGCGCTG 2124
Db 621 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
QY 2125 CAGGACGCGGACGAGGTGACATCGTACCGACGACGACGACGACGACGACGACGACG 2184
Db 641 GlnAspSerGlyLeuGluValAlaIleValThrAspSerGlnTyraAlaLeuGlyIleIle 660
QY 2185 CAGCGCCAGCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2244
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIle 680
QY 2245 AAGAAGAGAGGTGTACTGAGCTGGTGGTCCCGCCCGCCACGAGGATCGCGCGCAAGAG 2304
Db 681 LysLysGlnLysValTyraLeuAlaTrpValProAlaHisLysGlyIleGlyAsnGlu 700
QY 2305 CAGATCGACAGCTGGTGACAGGAGCATCCGCAAGGTGCTGTCTCGACGCGCATCGAT 2364
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 13
US-07-743-357-8
; Sequence 8, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: RF
; US-07-743-357-8

Alignment Scores:
Pred. No.: 3,93e-206 Length: 1003
Score: 3456.50 Matches: 647
Percent Similarity: 95.28% Conservative: 39
Best Local Similarity: 89.86% Mismatches: 29
Query Match: 75.58% Indels: 5
DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x US-07-743-357-8 (1-1003)
QY 220 TTCTTCGGGAGGACCTGGCTTCCCGAGGCGAGCGCGAGTCCCGAGGAGCAG 279
Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluLeuSerSerGluGln 20
QY 280 AACCGCGCGCAACAGCCCGACCGAGCTGCAGGTGCGCGC---GACAAACCCCGC 336
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSerLeu 40
QY 337 AGCGAGCGCGCGCGAGCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
Db 41 SerGluAlaGlyLysAspArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60
QY 391 TGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 450
Db 61 TrpGlnArgProIleValThrValLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80
QY 451 ACCGCGCGCGCGAGCACACCGCTGCTGGAGGAGATGAGCTGCCCGGCAAGTGGAGCCCAAG 510
Db 81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100
QY 511 ATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
Db 101 MetIleGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleLeuLeuGlu 120
QY 571 ATCTGGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
Db 121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140
QY 631 GGCGCGCAACATGCTGACCCAGCTGGCGCTGCACCTGCACTTCCCGCATCAGCCCCATCGAG 690
Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160
QY 691 ACCGTCCCGGTGAAGTGAAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180
QY 751 ACCGAGGAGAGAGTGAAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810
Db 181 ThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLys 200
```



;  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: SF2  
; US-07-743-357-7

Alignment Scores:  
Pred. No.: 1.15e-205 Length: 1004  
Score: 3449.00 Matches: 646  
Percent Similarity: 94.87% Conservative: 38  
Best Local Similarity: 89.60% Mismatches: 31  
Query Match: 75.42% Indels: 6  
Dbs: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x US-07-743-357-7 (1-1004)

QY	220	TTCTTCGGGAGGACCTGGCTTCCCGCCAGGCGAGCCCGCGAGTTCCTCCAGCGAGCAG	279
DB	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
QY	280	AACCGGCGCAACAGCCGCCACCGCGGAGCTCAGTGGCGCGC-----GACAAACCCC	333
DB	21	ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer	40
QY	334	CGCAGCGAGCGCGCGCGAGCGCGAGCGCAGGCCACCTG-----AACTTCCCGCAGATCAC	387
DB	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr	60
QY	388	CTGTGGCAGCGCCCTGTGTGACATCAAGTGGCGGCGCAGATCAAGGAGGCCCTGCTG	447
DB	61	LeuTrpGlnArgProLeuValThrIleArgIleGlyGlnLeuLysGluAlaLeuLeu	80
QY	448	GACACCGCGCGCGACACCGTGTGGAGGAGATGAGCTGCGCGCGAGTGGAGGCC	507
DB	81	AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro	100
QY	508	AGATGATCGCGCGCGCTGCGGCTTCATCAAGTGGCGGCGAGTACGACGAGCTCTGATC	567
DB	101	LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProVal	120
QY	568	GAGATCTGCGCGCAAGAGCGCATCGGCACCTGCTGATCGCGCGCCACCGCGTGAACATC	627
DB	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
QY	628	ATCGCGCGCAACATGTCGACCCAGTGGTGGCTGACCTGAACTTCCCATCAGCCCATC	687
DB	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
QY	688	GAGACCGTGCCTGAGCTGAAGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCG	747
DB	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
QY	748	CTGACCGAGGAGAGATCAAGCGCCCTGACCGCCATCTCGCAGGAGATGGAGAGAGGCG	807
DB	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
QY	808	AGATACCAAGATCGCGCGCGCGAGACCCCTACAAACCCCGCTGTTCGCATCAAGAG	867
DB	201	LysIleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys	220
QY	868	AAGCAGACCAAGTGGCGCAAGCTGTGGACTTCGCGAGCTGAACAGCGCCACCCAG	927
DB	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
QY	928	GACTTCTGGAGGTGACGTGGGCGCATCCCGACCGCGCGCGCGCGCGCGCGCGCGCG	987
DB	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
QY	988	GTGACCGTGTGAGCTGGGCGAGCGCTACTTCAAGTGGCTGGCGAGGAGCTTCCGC	1047
DB	261	ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheArg	280

QY	1048	AAGTACACCGCCTTACCATCCCGAGATCAACACGAGACCCCGCGCATCCGCTACCAG	1107
DB	281	LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln	300
QY	1108	TACAACTGCTGCTCCCGAGGCTGGAGAGGCGACCCCGAGCATCTTCAGAGCAGCATGACC	1167
DB	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320
QY	1168	AAGATCTCGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGTATCTACAG-----GCC	1221
DB	321	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyGlnTyMetAsp	340
QY	1222	CCCTCTGATCGTGGCAGCAGCATCGAGATCGGCGCAGCACCGCGCGCAGAGTTCGAGGAGCTG	1281
DB	341	AspLeuTyThrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
QY	1282	CGCAAGCAGCTGTGCGTGGGCTTCCACACCCCGCGCAAGACGAGACCCAGAGGAGGCC	1341
DB	361	ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
QY	1342	CCCTTCTGTGATGGCTACGAGCTGCACCCGCAAGTGGACCTGCAGCCCATCGAG	1401
DB	381	ProPheLeuTrpMetGlyTyThrGluLeuHisProAspLysTrpThrValGlnProIleMet	400
QY	1402	CTGCCCGAGAGGAGAGCTGGACCGCTGAACGACATCCAGAAGCTGGTGGCGCAAGCTGAAC	1461
DB	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
QY	1462	TGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGCAGCTGTGAAGCTGTGCGCGGC	1521
DB	421	TrpAlaSerGlnIleTyAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly	440
QY	1522	GCCAGGCGCTGACGACATCGTCCCTGAGCGGAGGCGCGAGCTGGAGCTGGCGCGAG	1581
DB	441	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuAlaGlu	460
QY	1582	AACCGCGAGATCTCGCGAGCGCGCTGTCACGCGCGTGTACTACGACCCCGAGCAGGACCTG	1641
DB	461	AsnArgGluIleLeuLysGluProValHisGluValTyThrAspProSerLysAspLeu	480
QY	1642	GTGCGCGAGATCCAGAAGCAGGCGCACGACCTGAGTGGACCTTACAGATCTTACAGGAGCCC	1701
DB	481	ValAlaGluIleGlnLysGlnGlyGlnTyThrTrpThrTyThrGlnIleTyThrGlnGluPro	500
QY	1702	TTCAGAACTGAAGACCGCAAGTACGCAAGATCGCGACCGCGCGCGCGCGCGCGCGCG	1761
DB	501	PheLysAsnLeuLysThrGlyLysTyAlaArgMetArgGlyAlaHisThrAsnAspVal	520
QY	1762	AAGCAGCTGACGAGCGCGCTGCAGAAAGATCGCCCATGGAGAGCATCGTGTCTGGGCGAAG	1821
DB	521	LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys	540
QY	1822	ACCCCGAAGTTCGCTGCGCTCCAGAGGAGACCTGGGAGACCTGGGAGCGCGCGCGCTAC	1881
DB	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpMetGluTy	560
QY	1882	TGGCGCGCACCTGGATCCCGAGTGGAGTTCGTGAACACACCCCGCGCGCGCGCGCGCTG	1941
DB	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu	580
QY	1942	TGTTACAGCTGGAGAGGAGCCCATCATCGGCGCGCGAGACCTTCTACGTTGAGCGCGCC	2001
DB	581	TrpTyThrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyThrValAspGlyAla	600
QY	2002	GCCAAACCGGAGACCAAGATTCGGCAAGCGCGGTCTAGTACCGACCGCGCGCGCGCGAG	2061
DB	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyThrValThrAspArgGlyArgGlnLys	620
QY	2062	ATCGTCAGCTGACCGCAGACCCCAACAGAGACCGAGCTGCAGCGCATCTCCAGCTGGCC	2121
DB	621	ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla	640
QY	2122	CTGCAGGACAGCGCGCAGGAGTGAACATCGTGAACCGCAGCAGCGCATCGCCCTGGGCATC	2181

```

Db      641 LeuGlnaspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Qy      2182 ATCCAGGCCAGCCGCAAGAGGAGGAGCGAGCTGGTGAACACCATCATCGAGCAGCTG 2241
Db      661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
Qy      2242 ATCAAGAGGAGAGAGGTGTACTGAGCTGAGTGGTGCCCGCCCAAGAGGAGCATCGGCGGCAAC 2301
Db      681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
Qy      2302 GAGCAGATCGACAGCTGGTGAAGGAGGATCGGAGGAGTGTCTTCCTGGAGCGGCATC 2361
Db      701 GlnValAlaAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720
Qy      2362 GAT 2364
Db      721 Asp 721

RESULT 15
US-09-319-588C-6
; Sequence 6, Application US/09319588C
; Patent No. 6509018
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAULIERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSI, Francoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/09/319,588C
; CURRENT FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-319-588C-6

Alignment Scores:
Pred. No.:      3,17e-189      Length:      1014
Score:          3183.50      Matches:      583
Percent Similarity: 90.25%      Conservative: 74
Best Local Similarity: 80.08%      Mismatches: 58
Query Match:      69.62%      Indels:      13
DB:              4          Gaps:      4

US-09-610-313B-31 (1-2463) x US-09-319-588C-6 (1-1014)
Qy      220 TTCTTCGCGAGGACCTTCCCGCGGCAAGCGCCGAGTTCCTCCGAGCGAGCAG 279
Db      1 PhePheargGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProAspAsn 20
Qy      280 AAC-----CGCGCAACAGCCCAACAGCGCGCGAGCTGCAGGTG----- 318
Db      21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40
Qy      319 ---CGCGGCGACACCCCGCAGCGCGCGCGCGCGCGCCAG-----GGCACC 366
Db      41 HisThrGlyGluGlyAspAlaGlyProGlyGluAspArgGluLeuSerValProThr 60
Qy      367 CTGAACCTTCCCGCAGATCACCCTGTGGCAGCGCGCCCTGGTGAGCATCAAGGTGGCGGC 426
Db      61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
Qy      427 CAGATCAAGGAGGCCCTGTGGACACCGCGCGCGCGACACCGTGTGGAGGAGATGAGC 486

```

```

Db      81 GluValargGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluLeuGln 100
Qy      487 CTGCGCCGCGAAGTGCAGAACCAAGATCATCGCGGCATCGCGGCTTCATCAAGGTGGCG 546
Db      101 LeuGluGlyLysTrpLysProLysMetIleGlyIleGlyGlyPheIleLysValarg 120
Qy      547 CAGTACGACCATCTCTGATCGAGATCTGCGGCAAGAGCCATCGGCACCGTGTGTATC 606
Db      121 GlnTyrAspAsnIleThrValAspIleGlnArgLysAlaValGlyThrValLeuVal 140
Qy      607 GCGCCCGACCCCGTGAACATCATCGCGGCAACATGCTGACCCAGCTGGGCTGCACCCGTG 666
Db      141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160
Qy      667 AACTTCCCATCAGCCCATCGAGACCGTGCCTGAAGCTGAAGCCCGCGCATGACACGCG 726
Db      161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180
Qy      727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTTGACCGCCATCTGC 786
Db      181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200
Qy      787 GAGGAGATGGAGAGGAGGCGCAGATCACCAAGATCGCGCGGAGAACCCCTACACACC 846
Db      201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220
Qy      847 CCGGTGTTCCCATCAAGAAGAAGACAGACAGCAACCAAGTGGCGCAAGCTGCTGCATCTCCGC 906
Db      221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPhearg 240
Qy      907 GAGCTGAACAGCCACCCAGGACTTCTGGAGGTGCAGTGGCATCCCCACCCCGCC 966
Db      241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260
Qy      967 GGCTGAAGAAGAAGAGCGTGACCGTGTGACGTGGCGCGCGCTACTTCTCAGCGTG 1026
Db      261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280
Qy      1027 CCCCTGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAACAACGAG 1086
Db      281 ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu 300
Qy      1087 ACCCGCGCATCCGCTACCAAGTACACGTGCTGCCCGAGGCTGGAAGGCGAGCCCCAGC 1146
Db      301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320
Qy      1147 ATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAACCCCGAGTTC 1206
Db      321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340
Qy      1207 GTGATCTACCAG-----GCCCGCTGTAGTGGCGAGCGACCTGGAGATCGGCCAGCAC 1260
Db      341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360
Qy      1261 CGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGTGCGCTGGGGCTTCACCAACCCCGAC 1320
Db      361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380
Qy      1321 AAGAGCACCAGAGAGGAGGAGCCCTTCTGTGGATGGGTACGAGCTGCACCCCGCAGCAG 1380
Db      381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400
Qy      1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db      401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420
Qy      1441 AAGTGTGTGGCGCAAGCTGAACCTGGCGCAGCAGATCTACCCCGCATCAAGGTGGCGCAG 1500
Db      421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440
Qy      1501 CTGTGCAAGCTGTGCGCGCGCCCAAGCGCTGACCGACATCGTGCCCTGACCGGAGAG 1560

```



```

Db      441  LeuCysLeuLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460
QY      1561  GCCGAGCTGGAGCTGGCCGAGAACCGGAGATCCTGCGGAGCCCGTGACACGCGCTGTAC 1620
Db      461  AlAGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480
QY      1621  TACGACCCCAAGCAAGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAAGTGGACC 1680
Db      481  TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnTyrThr 500
QY      1681  TACGAGATCTACGAGAGCCCTTCAGAACTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db      501  TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520
QY      1741  ACCGCCCAACCAACGAGTGAAGAGCTGACCGAGCGCGTGCAGAAAGATCGCCATGGAG 1800
Db      521  SerAlaHisThrAsnAspIleLysGlnLeuValGluValArgLysValAlaThrGlu 540
QY      1801  AGCATCGTGATCTGGGCGAAGACCCCAAGTTCGGCTGCCCATCCAGAGGAGACCTGG 1860
Db      541  SerIleValIleTyrGlyLysThrProLysPheArgLeuProValGlnLysGluValTyr 560
QY      1861  GAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1920
Db      561  GluAlaTyrThrAspHisThrGlnAlaThrTyrPheValAsn 580
QY      1921  ACCCCCTGCTGAGCTGGTACCTGAGTGGAGAGAGAGCCCATCATCGGCGCGGAG 1980
Db      581  ThrProLeuValLysLeuTyrGlnLeuGluThrGluProLysSerGlyAlaGlu 600
QY      1981  ACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040
Db      601  ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620
QY      2041  ACCGACCGGGCGCGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Db      621  ThrAspArgGlyArgGlnLysValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640
QY      2101  CTCGAGGCCATCCAGCTGGCCCTGCAGGACAGCGGAGGAGTGAACATCGTGACCGAC 2160
Db      641  LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660
QY      2161  AGCCAGTACCGCTGGGCGATCATCCAGGCCCGCAGCCGACAAAGACGAGAGCGAGCTGGT 2220
Db      661  SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680
QY      2221  AACGAGATCATCGAGAGCTGATCAGAGAGAGAGTGTACCTGAGCTGGTGGCCGCC 2280
Db      681  SerGlnIleIleGluLeuIleLysLysGluArgValTyrLeuSerTyrValProAla 700
QY      2281  CACAAGGGCATCGCGCAACGAGCAGATCGCAAGCTGGTGAAGCAAGGGCATCCGCAAG 2340
Db      701  HisLysGlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720
QY      2341  GTGCTGTTCCTGGACGGCATCGAT 2364
Db      721  IleLeuPheLeuAspGlyIleGlu 728

```

Search completed: June 2, 2005, 05:12:37  
Job time : 109 secs

GenCoré version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 03:13:23 ; Search time 33.9172 Seconds  
(without alignments)  
10815.332 Million cell updates/sec

Title: US-09-610-313B-32  
Perfect score: 4555  
Sequence: 1 gtgcagccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model  
-Q=/cgn2\_1/USPTO/US09610313/runat\_31052005\_155137\_151173/app\_query.fasta\_1.7893  
-DB=Issued\_Patents\_AA -QMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CIGN 1 101 @runat\_31052005\_155137\_151173 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pcp:\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pcp:\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pcp:\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pcp:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pcp:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	3454	75.8	1005	2	US-07-743-357-1
2	3450.5	75.8	1003	2	US-07-743-357-9
3	3432	75.3	1015	3	US-08-463-210-9
4	3432	75.3	1015	3	US-09-124-900-3
5	3432	75.3	1015	4	US-08-463-028-9
6	3432	75.3	1016	2	US-07-743-357-2
7	3429	75.3	1003	4	US-09-309-572-17
8	3429	75.3	1003	4	US-09-718-096-17
9	3426	75.2	1016	2	US-07-743-357-5
10	3425	75.2	1016	2	US-07-743-357-3
11	3422	75.1	1016	2	US-07-743-357-4
12	3416.5	75.0	1003	2	US-07-743-357-10

## SUMMARIES

## ALIGNMENTS

## RESULT 1

US-07-743-357-1  
; Sequence 1, Application US/07743357  
; Patent No. 5858646

## GENERAL INFORMATION:

; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological

; ACTIVITY OF INVENTION: activity for use as diagnostic reagent and/or vaccine

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

ADDRESSEE: KIRBY EADES GALE BAKER

STREET: Box 3432, Station D

CITY: Ottawa

STATE: Ontario

COUNTRY: Canada

ZIP: K1M 1H8

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/07743,357

APPLICATION NUMBER: US/07743,357

FILING DATE: 21-AUG-1991

CLASSIFICATION: 424

PRIOR APPLICATION DATA: PCT/CA90/00062

APPLICATION NUMBER: 23-FEB-1990

FILING DATE: 23-FEB-1990

NAME: Gale, Edwin J.

REGISTRATION NUMBER: 28,584

REFERENCE/DOCKET NUMBER: 30924-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (613) 237-6900

TELEFAX: (613) 237-0045

INFORMATION FOR SEQ ID NO: 1:

Sequence 8, Appli  
Sequence 7, Appli  
Sequence 6, Appli  
Sequence 22, Appl  
Sequence 35, Appl  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 2, Appli  
Sequence 8, Appli  
Sequence 4, Appli  
Sequence 14, Appl  
Sequence 14, Appl  
Sequence 1, Appli  
Sequence 46, Appl  
Sequence 24, Appl  
Sequence 16, Appl  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Patent No. 5223423  
Patent No. 5223423  
Sequence 1, Appli  
Sequence 72, Appl  
Sequence 71, Appl  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Patent No. 5252477  
Patent No. 5252477  
Sequence 9, Appli  
Sequence 9, Appli  
Patent No. 5320958

13 3413.5 74.9 1003 2 US-07-743-357-8  
14 3406 74.8 1004 2 US-07-743-357-7  
15 3140.5 68.9 1014 4 US-09-319-588C-6  
16 3098 68.0 913 2 US-07-743-357-22  
17 3076.5 67.5 1350 4 US-09-952-060-35  
18 3035 66.6 913 2 US-07-743-357-6  
19 2779.5 61.0 875 4 US-09-952-060-6  
20 2774 60.9 850 4 US-09-952-060-2  
21 2755.5 60.5 875 4 US-09-952-060-8  
22 2750 60.4 850 4 US-09-952-060-4  
23 2746 60.3 562 3 US-09-117-217-14  
24 2746 60.3 562 4 US-09-735-487-14  
25 2715 59.6 560 4 US-10-205-641-1  
26 2464.5 54.1 1018 4 US-09-206-551-46  
27 2445.5 53.7 1016 4 US-09-625-972-24  
28 2216 48.6 441 4 US-09-991-358-16  
29 2173 47.7 1055 2 US-08-659-251-5  
30 2173 47.7 1055 3 US-09-256-490-5  
31 2173 47.7 1055 5 PCT-US96-11445-5  
32 2154.5 47.3 3080 6 5223423-4  
33 2154.5 47.3 3080 6 5223423-4  
34 2111 46.3 427 3 US-09-690-265-1  
35 1529 33.6 327 3 US-08-679-493A-72  
36 1512 33.2 327 3 US-08-679-493A-71  
37 1509 33.1 314 1 US-08-589-446-6  
38 1509 33.1 314 1 US-08-444-882-6  
39 1509 33.1 314 2 US-08-389-459A-6  
40 1509 33.1 314 3 US-08-987-867A-6  
41 1309 28.7 275 6 5252477-3  
42 1309 28.7 275 6 5252477-3  
43 1272 27.9 1150 3 US-09-238-303-9  
44 1272 27.9 1150 4 US-09-946-239-9  
45 1254 27.5 261 6 5320958-3

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1005 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
; US-07-743-357-1

Alignment Scores:
Pred. No.: 9,89e-208 Length: 1005
Score: 3454.00 Matches: 651
Percent Similarity: 95.01% Conservative: 34
Best Local Similarity: 90.29% Mismatches: 28
Query Match: 75.83% Indels: 8
DB: 2 Gaps: 4

US-09-610-313b-32 (1-2457) x US-07-743-357-1 (1-1005)

QY 220 TTCTTCCGCGAGACTGCGCTTCCCGAGGCAAGGCGCGAGTTCCTCCAGCGAGCAG 279
Db 2 PhePheAsgGluAaspLeuAlaPheLeuGlnGlyLysAlaAargGluPheSerSerGluGln 21
QY 280 AACCGCGCAACAGCCGACACAGCCGAGCTGCAGGTG-----CGCGGCGCAACACCC 333
Db 22 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 41
QY 334 CGCAGCGAGGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
Db 42 ProSerGluAlaGlyAlaAaspArgGlnGlyThrValSerPheAsnPheProGlnValThr 61
QY 388 CTGTGGCAGCGCCCTGTGTAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
Db 62 LeuTrpGlnAargProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 81
QY 448 GACACCGCGCGCGACACACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAGCGCC 507
Db 82 AspThrGlyAlaAaspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 101
QY 508 AGATGATCGCGCGCATCGCGGCTTCATCAGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 567
Db 102 LysMetIleGlyGlyIleGlyPheIleLysValAargGlnTyrAspGlnIleLeuIle 121
QY 568 GAGATCTGCGGCAAGAGCCATCGCGACCGCTGCTGATCGGCGCGCGCGCGCGCGCGCG 627
Db 122 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 141
QY 628 ATCGCGCGCAACATGTCACCGAGTGGGTGTCACCTGAACTTCCCATCAGCGCCCATC 687
Db 142 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 161
QY 688 GAGACGTCGCGTGAAGTGAAGCCGCGCATGGAGCGCGCGCGCGCGCGCGCGCGCGCG 747
Db 162 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 181
QY 748 CTGACCGAGGAGAAAGATCAAGCGCTGACCGCGCATCTGCGAGAGATGGAGAGAGGCG 807
Db 182 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetLysGluGly 201
QY 808 AAGATCAACAAGATCGGCGCGCGAGAACCCCTCAACACCCCGTGTTCGCCATCAAGAAG 867
Db 202 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 221
QY 868 AAGGACAGCAACAAGTGGCGAGTGGTGGACTTCCGGAGCTGAACAGCGCGCACCCAG 927
Db 222 LysAspSerThrLysTrpArgLysLeuValAspPheAargGluLeuAsnLysArgThrGln 241
QY 928 GACTTCTGGAGGTGAGCTGGGCGATCCCCACCGCGCGCGCGCGCGCGCGCGCGCGCG 987

```

```

Db 242 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 261
QY 988 GTGACCGTGTGGAGCGTGGGCGAGCGCTACTTACGCGTGGCCCTCCGAGGAGACTTCGCG 1047
Db 262 ValThrValLeuAaspValGlyAaspAlaTyrPheSerValProLeuAaspGluAaspPheArg 281
QY 1048 AAGTACACCGCTTACCATCCCGAGCATCAACACGAGACCCCGCGGATCCGCTACACAG 1107
Db 282 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 301
QY 1108 TACAACGTGTGCGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167
Db 302 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 321
QY 1168 AAGATCTTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1221
Db 322 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 341
QY 1222 CCCCTGTACGTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCGCAAGATCGAGGAGCTG 1281
Db 342 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 361
QY 1282 CGCAAGCACCTGTGCGTGGGCTTACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1341
Db 362 ArgGlnHisLeuLeuAargTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 381
QY 1342 CCTTCTCTG-----CCATCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1395
Db 382 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 401
QY 1396 CTGCGCGAGAGAGAGAGCTGGACCGTGAACACACATCCAGAAAGCTGGTGGCAAGCTGAAC 1455
Db 402 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 421
QY 1456 TGGGCGACGACAGATCTACCGCGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCG 1515
Db 422 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuAargGly 441
QY 1516 GCCAAGCGCTGACGAGCATCGTCCCTGACCGAGGCGCGAGCTGGAGCTGGCGCGCGAG 1575
Db 442 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 461
QY 1576 AACCGCGAGATCTGCGCGAGCGCGTGCACCGCGTGTACTACGACCGCGCGCGCGCGCG 1635
Db 462 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 481
QY 1636 GTGCGCGAGATCCAGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1695
Db 482 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 501
QY 1696 TTCAGAAGACCTGAAGACCGCGCAAGTACGCCAAGATGGCGCGCGCGCGCGCGCGCGCG 1755
Db 502 PheLysAsnLeuLysThrGlyLysTyrAlaAargMetArgGlyAlaHisThrAsnAspVal 521
QY 1756 AAGCAGCTGACCGAGGCGCGTGCAGAAGATCCCATGGAGAGCATCGTGTATCTGGGCGCAAG 1815
Db 522 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 541
QY 1816 ACCCGCGAGTTCGCGCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db 542 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 561
QY 1876 TGGCAGGCGCACCTGATCCCGAGTGGAGTTCGTGAACACACCGCGCGCGCGCGCGCGCG 1935
Db 562 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 581
QY 1936 TGGTACCACTGAGGAGAGAGCGCGCATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1995
Db 582 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 601
QY 1996 GCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCCCGGCGCGCGAGAG 2055
Db 602 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 621

```



Db 301 AenValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 QY 1171 ATCTGGAGCCCTTCGGCCCGCAACCCCGAGATCGTATACAG-----CCCCC 1224  
 Db 321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
 QY 1225 CTGTACGTGGGAGCGACCTGGAGATCGCCAGACCGCGCCCAAGATCGAGGAGCTGGC 1284  
 Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuArg 360  
 QY 1285 AAGCACTCTGCTGGCTGGCTTACACACCCCGCAAGACCAAGAGAGGCCCCC 1344  
 Db 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1345 TTCCTG-----CCCATCGAGCTGCACCCCGCAAGTGCACCGTCCAGCCATCAGCTG 1398  
 Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
 QY 1399 CCGGAGAAGGAGCTGGACCGTGAACACATCCAGAACTGTGGGCAAGCTGAACCTGG 1458  
 Db 401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
 QY 1459 GCCAGCCAGATCTACCCCGCATCAAGTGGCCAGCTGTGCAAGCTCTGCGCGGCC 1518  
 Db 421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440  
 QY 1519 AAGGCCCTGACGATCTGCTCCCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAAC 1578  
 Db 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuLeuAlaGluAsn 460  
 QY 1579 CCGAGATCTCTCGGAGCGCTGCACGCGGTACTAGCAGCCCGCAGCAAGGAGCTGTG 1638  
 Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
 QY 1639 GCCAGATCTCAGAAGCAGCGCCACAGCAGTGGACCTACCATCTACGAGGAGCCCTTC 1698  
 Db 481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGlnTyr 500  
 QY 1699 AAGACCTGAGACCGGCAAGTACCGCAAGATCGCCAGCGCCACCAACCAAGCTGAG 1758  
 Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
 QY 1759 CAGCTGACCGAGCGCTGCAGAAAGTCCGATGAGAGATCGTGTATCTGGGCGAAGACC 1818  
 Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
 QY 1819 CCCAGTTCCTGCTCCCATCCAGAGAGACCTGGAGACCTGTGGACCGACTACTGG 1878  
 Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560  
 QY 1879 CAGGCCACTGTATCCCGAGTGGAGTTCGTGAACACCCCGCTCGTGAAGCTGTGG 1938  
 Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580  
 QY 1939 TACCACTGGAAGAGCGCCATCATCGCGCGAGACCTTCTAGTGGAGCGCGCCGCC 1998  
 Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 QY 1999 AACCGCGAGACCAAGATCGCAAGCGCGCTACGTACCGACCGACCGCGCGCGAGATC 2058  
 Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgLysArgGlnLysVal 620  
 QY 2059 GTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCAGCTGCCCTG 2118  
 Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
 QY 2119 CAGGACAGCGGAGGTGAACATCGTACCGCAGCAGTACGCTCGCTGGGATCATC 2178  
 Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 QY 2179 CAGGCCAGCCGACAGAGCGAGCGAGCTGGTGAACCAAGATCATCAGAGCTGATC 2238  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

QY 2239 AAGAAGGAGAAGGTGTACTGAGCTGGTGGCCGCCCAAGGGCATCGGCGGCAACGAG 2298  
 Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysGlu 700  
 QY 2299 CAGATCCGACAAAGCTGGTGAGCAAGGCAATCCGCAAGGTGTCTTCTCGACGCGCATCGAT 2358  
 Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720  
 RESULT 3  
 US-08-463-210-9  
 ; Sequence 9, Application US/08463210  
 ; Patent No. 6001977  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHANG, Nancy T.  
 ; APPLICANT: GALLO, Robert C.  
 ; APPLICANT: WONG-STAAAL, Flossie  
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morgan & Finnegan, L.L.P.  
 ; STREET: 345 Park Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10154-0053  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/463.210  
 ; FILING DATE: 05-JUN-1995  
 ; CLASSIFICATION: 436  
 ; PRIOR APPLICATION NUMBER: US 06/693,866  
 ; FILING DATE: 23-JAN-1985  
 ; APPLICATION DATA:  
 ; APPLICATION NUMBER: US 06/659,339  
 ; FILING DATE: 10-OCT-1984  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Serunian, Leslie A.  
 ; REGISTRATION NUMBER: 35,353  
 ; REFERENCE/DOCKET NUMBER: 2026-4193US2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 758-4800  
 ; TELEFAX: (212) 751-6849  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1015 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: HTLV-III  
 ; FEATURE:  
 ; NAME/KEY: Protein  
 ; LOCATION: 1..1015  
 ; OTHER INFORMATION: /note= "pol protein of HTLV-III"  
 US-08-463-210-9  
 Alignment Scores:  
 Pred. No.: 2,346-206 Length: 1015  
 Score: 3432.00 Matches: 650  
 Percent Similarity: 93.18% Conservative: 33  
 Best Local Similarity: 88.68% Mismatches: 20  
 Query Match: 75.35% Indels: 20  
 DB: 3 Gaps: 5  
 US-09-610-313B-32 (1-2457) x US-08-463-210-9 (1-1015)

QY 220 TTCTTCCGCGAGGACCTGGCTTCCCTCCCGAGGCAAGGCGCCGAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 268 -----CCGAGCGAGCAGAACCGCGCCACACAGCCGCCACCGC 303  
Db 21 ThrArgAlaAenSerProThrIleSerSerGluGlnThrArgAlaAenSerProThrArg 40  
QY 304 CGCGAGCTGCAGGTG-----CGCGCGGACAAACCCCGCAGCGCGCGCGCGCGCGC 357  
Db 41 ArgGluLeuGlnValTyrGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
QY 358 CAGGCGACCTCG-----AATCTCCCGCAGATACCTCTGGCGAGCGCCCTCGGTGAGC 411  
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTyrGlnArgProLeuValThr 80  
QY 412 ATCAAGGTGGCGCGCCAGATCAAGGAGCGCTGTGGACACCGCGCGCGCGCGCACCGTG 471  
Db 81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuVal 100  
QY 472 CTGAGGAGATGAGCTCGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531  
Db 101 LeuGluMetSerLeuProGlyArgTyrPylsProLysMetIleGlyGlyGlyGly 120  
QY 532 TTCTATCAAGGTGCGCCAGTACGACAGATCTGATCGAGATCTCGCGCAAGAGCCCATC 591  
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
QY 592 GGCACCTGTGTGTCGCGCGCCCGCCCGTGGACATCATCGCGCGCCCAATGCTGACCGCAG 651  
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuGln 160  
QY 652 CTGGGCTGCACCTCGAATCTCCCATCAGCCCATCGAGCGCGCGCGCGCGCGCGCGCGCG 711  
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CCGCGCATGAGCGCCCGCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCC 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTyrProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCGCCCATCTCGGAGGAGATGAGAGAGGCGCAAGATCACCAGATCGCGCGCGCAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleLysIleGlyProGlu 220  
QY 832 AACCCCTACAAACCCCGCTGTTCGCCATCAAGAAGAGGACACACCAAGTGGCGCAAG 891  
Db 221 AsnProLysAsnThrProValPheAlaIleLysLysLysAspSerThrLysTyrArgLys 240  
QY 892 CTGTTGGAATTCGCGAGCTGAACAAGCGCACCGCAGGACTTCTGGGAGGTGCAGCTGGGC 951  
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTyrGluValGlnLeuGly 260  
QY 952 ATCCCCCACCAGC 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTTCAGGTGCG 1071  
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACACAGAGACCCCGCGCATCCGCTACAGTACCAACGTGTGCGCGCGCGCGCGCG 1131  
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyr 320  
QY 1132 AAGGCGAGCG 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCCGAGATCGGTACTACCG-----GCCCGCTGTACGTGGCGCGCGCGCGCG 1245  
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

QY 1246 GAGATCGCGCGCACCGCGCGCAAGATCGAGAGCTGGCGCAAGCACCTCTGCTGCTGGGCGC 1305  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTyrGly 380  
QY 1306 TTCACACCCCGCGACAGACAGACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1359  
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTyrMetGlyTyrGlu 400  
QY 1360 CTGCACCCCGCACCAAGTGGACCGTCCAGTCCAGTGGCGCGCGCGCGCGCGCGCGCGCG 1419  
Db 401 LeuHisProAspLysTyrThrValGlnProIleValLeuProGluLysAspSerTyrThr 420  
QY 1420 GTGAACACGATCCAGAGCTGGTGGCAAGCTGAACCTGGCGCGCGCGCGCGCGCGCGCG 1479  
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTyrPalaSerGlnIleTyrProGly 440  
QY 1480 ATCAAGGTGGCG 1539  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1540 CCCCTGACCGAGGAGCG 1599  
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1600 GTGCACCGCTGTACTACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1659  
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1660 CACGACGAGTGGACCTACCGATCTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1719  
Db 501 GlnGlyGlnTyrThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1720 TAGCCCAAGATGGCG 1779  
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1780 AAGATCGCCATGGAGAGCATCGTATCTGGGCGCAAGACCCCGCGCGCGCGCGCGCGCG 1839  
Db 541 LysIleThrThrGluSerIleValIleTyrPylsThrProLysPheLysLeuProIle 560  
QY 1840 CAGAGGAGACCTGGGAGACCTGTGGACCGACTACTGGCGCGCGCGCGCGCGCGCGCGCG 1899  
Db 561 GlnLysGluThrTyrGluThrTyrTyrThrGluTyrTyrGlnAlaThrTyrPylProGlu 580  
QY 1900 TGGGAGTTCGTGAACCCCG 1959  
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTyrPylGlnLeuGluLysGluPro 600  
QY 1960 ATCATCG 2019  
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2020 AAGCGCGCTACGTGACCG 2079  
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2080 AACGAGAGACCGAGCTGCG 2139  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2140 AACATCGTGAACGACCG 2199  
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2200 GAGAGCGAGCTGTGAACCGAGATCATCGACGCTGATCAAGAGGAGAGAGGTGTACTGT 2259  
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700  
QY 2260 AGCTGGTGTCCCG 2319  
Db 701 AlaTyrValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2320 AAGGCGATCGCGAAGGTGTCTTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2358



```
Db      721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
|||||
RESULT 4
US-09-124-900-3
; Sequence 3, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andrea
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: PREDL, Renate
; APPLICANT: SCHWATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112p
; CURRENT APPLICATION NUMBER: US/09/124,900
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: PCT/EP95/01481
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-124-900-3

Alignment Scores:
Pred. No.:      2,346-206      Length:      1015
Score:          3432.00      Matches:      650
Percent Similarity: 93.18%      Conservative: 33
Best Local Similarity: 88.68%      Mismatches: 30
Query Match:      75.35%      Indels:      20
Db:              3      Gaps:      5

US-09-610-313B-32 (1-2457) x US-09-124-900-3 (1-1015)
QY      220 TTCTTCCGGAGGACTGGCTTCCCGAGGCGAGGCGCGAGTTC----- 267
Db      1 PhePheArgLysLeuAlaPheLeuGlnGlyLysAlaArgLysPheSerSerGluGln 20
QY      268 -----CCGAGCGAGCAGAACCGCGCCAAACAGCCCAACCCAGC 303
Db      21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY      304 CCGCAGCTGCAGTG-----CGCGGCGACAAACCCCGAGCGAGCGCGCGCGCGCGC 357
Db      41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY      358 CAGGCGACCCCTG-----AATCTCCCGACATCACCTCTGCGCAGCGCCCTGTGTGAGC 411
Db      61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY      412 ATCAAGGTGGGGCGGCACATCAAGAGGCGCTTGTGGACACCGCGCGCGCGACACCCGTG 471
Db      81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY      472 GTGGAGGAGATGAGCTGCCCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGCGC 531
Db      101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
QY      532 TTCATCAGGTGCGCGCAGTACACACAGATCCTGTATCGAGATCTGGCGCAAGAGGCCATC 591
Db      121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY      592 GGCACCGTGTGATCGGCGCCCAACCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAG 651
|||||
141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY      652 CTGGGCTGCACCCCTGAATCTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAG 711
Db      161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
QY      712 CCGGCGATGGACGGCCCAAGGTGAAGCAGTGGGCGCTTGCAGCGAGGAGAAAGATCAAGGCC 771
Db      181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY      772 CTGACCGCCATCTCGAGGAGATCGAGAGAGAGGCAAGATCACCAAGATCGGCCCGCAG 831
Db      201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
QY      832 AACCCCTACAACACCCCGCTGTTCGCCATCAAGAAAGAGGACAGCACCAAGTGGGCGCAG 891
Db      221 AsnProTyraAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240
QY      892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAGACTTCTGGGAGGTGCAGCTGGCG 951
Db      241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY      952 ATCCCCCAACCCCGCGCTGGAAGAAGAGAGCGTACCGTCTCGACGTGGCGCAG 1011
Db      261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280
QY      1012 GCCTACTTCAGGTGCGCCCTGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCC 1071
Db      281 AlaTyraPheSerValProLeuAspGluAspPheArgLysTyraAlaPheThrIlePro 300
QY      1072 AGCATCAACAACAGAGACCCCGCGCATCCGTACAGTACAACAGTCTGCTCCCGAGGCTGG 1131
Db      301 SerIleAsnAsnGluThrProGlyIleArgTyraGlnTyraAsnValLeuProGlnGlyTrp 320
QY      1132 AAGGCGACCCCGCAGCATCTTCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCC 1191
Db      321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
QY      1192 CGCAACCCCGAGATCGTGATCTACAG-----GCCCGCTGTAGTGGGCGAGCACCTG 1245
Db      341 GlnAsnProAspIleValIleTyraGlnTyraMetAspAspLeuTyraValGlySerAspLeu 360
QY      1246 GAGATCGGCGCAGCACCGCGCAAGATCGAGGAGCTCGCAGACCTCTGCTGGTGGCGC 1305
Db      361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
QY      1306 TTCACACCCCGCAGAACAGACACAGAGAGAGCCCGCTTCCTG-----CCCATCGAG 1359
Db      381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyraGlu 400
QY      1360 CTGCACCCCGCAGACAGTGCACCGCTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACC 1419
Db      401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY      1420 GTGAACGACATCCAGAAGCTGTGGCAAGCTGAATCGGCGCAGCAGATCTTACCCCGCGC 1479
Db      421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyraProGly 440
QY      1480 ATCAAGGTGGCGCCAGCTGTCAAGCTGTGCGCGCGCCAAAGCGCTTCAACGACATCGTG 1539
Db      441 IleLysValArgGlnLeuLysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY      1540 CCGCTGACCGAGGAGCGCGAGCTGGAGTGGCGCGAGAACCGCGAGATCTCGCGCGAGCC 1599
Db      461 ProLeuThrGluAlaGluLeuGluAlaGluAsnArgGluIleLysLysGluPro 480
QY      1600 GTGACGCGCTGTACTACGACCCCAAGCAAGCACTGTGTGGCGCGAGATCCAGAGCAGGCGC 1659
Db      481 ValHisGlyValTyraTyraProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY      1660 CACGACCGAGTGCATCTACAGATCTTACAGAGCGCTTCAAGAACCTTCAAGACCGCGCAG 1719
Db      501 GlnGlyGlnTrpThrTyraGlnIleTyraGlnGluProPheLysAsnLeuLysThrGlyLys 520
```

```
QY 1720 TACGCCAGATGCGCAGCCGCCACACCAACGACGTGAAGAGCTGACCGAGGCGGTGCG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1780 AAGATCGCCATGCGAGAGCATGCTGATCTGGGGCAAGACCCCAAGTTCGGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTrpGlyThrProLysPheLysLeuProIle 560
QY 1840 CAGAAGAGACCTGGGAGACCTGTGTGGACCACTACTGGCAGGCCACCTGATCCCGAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluThrTrpGlnAlaThrTripleProGlu 580
QY 1900 TGGAGTTCTGTGAACACACCCCGCTGTGAAGCTGTGTACCACTGGAGAGAGAGCC 1959
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTrpGlnLeuGluLysGluPro 600
QY 1960 ATCATCGCGCGGAGACCTTCTAGTGTGACCGCGCGCCCAACCGCGAGACCAAGATCGC 2019
Db 601 IleValGlyAlaGluThrPheTrpValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCGCGCTACGTGACCGACCGCGCGCGGCGGAGATCGTGAGCTGACCGAGACCAACC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACGAGAAGACCGAGCTGACGGCCATCCAGCTGCGCTGCGAGGACGCGGAGCGAGGTG 2139
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2140 AACATCGTGCGGACGACGACGACGCTGCGCTGCGGATCATCCAGGCGCCCGCAAGAGC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGACGAGCTGCTGAACACGATCATCGACGAGCTGATCAAGAGGAGAGGAGTGTACTG 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTGCGCGCCCAAGAGGATCGCGGCAACGAGCAGATCGACAAGCTGCTGAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyLysGlnGlnValAspLysLeuValSer 720
QY 2320 AAGGCGATCCCGACAGGTGCTCTTCTGACCGGATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 5
US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STRAAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
```

```
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; NAME/KEY: Protein
; LOCATION: 1..1015
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
; US-08-463-028-9

Alignment Scores:
Pred No.: 2,34e-206 Length: 1015
Score: 3432.00 Matches: 650
Percent Similarity: 93.18% Conservative: 33
Best Local Similarity: 88.68% Mismatches: 30
Query Match: 75.35% Indels: 20
DB: 4 Gaps: 5

US-09-610-313B-32 (1-2457) x US-08-463-028-9 (1-1015)
QY 220 TTCTTCGCGAGGACCTGGGCTTCCCCCAGGGCAAGGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 268 -----CCCGAGGAGAGAACCGCGCCCAACAGCCCGCCAGCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGCGACAAACCCCGCAGCAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGCGACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGTGGCGGCCAGATCAAGAGGCGCTGTGTGACACCGCGCGCGAGCAGCACCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTTCCCGCAACTGGAAGCCCAAGATGATCGCGCGGATCGCGGCG 531
Db 101 LeuGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyIleGly 120
QY 532 TTCATCAAGGTGGCCAGTACGACAGATCTGTATCGAGATCTCGCGCAAGAGGCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGGCGCCCGCCCGCGTGAACATCATCGCGCGCCCAACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGACCCCTGACTTCCCGCATCGCCCATCGAGCCCGCGCGCTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
```





```

Db      601  ilValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY      2020  AAGCGCGGTAGTACCGACCGCGCGCGGAGAGATCGTGTGACCGAGACCAACC 2079
Db      621  LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY      2080  AACCAAGAACACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGACGAGGTG 2139
Db      641  AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY      2140  AACATCGTGACCGACAGCAGTACGCCCTGGGCATCATCCAGGCCAGCCGACCAAGAGC 2199
Db      661  AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY      2200  GAGAGCGAGCTGTGAACAGATCATCGAGCAGCTGTATCAAGAAGGAGAGTGTACCTG 2259
Db      681  GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGlnLysValTyrLeu 700
QY      2260  AGCTGGTGGCCCGCCACAAAGGGCATCGCGCGCAAGCAGCAGATCGACAAGCTGTGAGC 2319
Db      701  AlaTrpValProAlaHisLysGlyIleGlyGlnGlnValAspLysLeuValSer 720
QY      2320  AAGGGCATCCGCAAGTGTCTGTCTTGGACGGCATCGAT 2358
Db      721  AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

```

RESULT 7

```

US-09-309-572-17
; Sequence 17, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-309-572-17

```

```

Alignment Scores:
Pred. No.:      3,6e-206      Length:      1003
Score:          3429.00      Matches:      648
Percent Similarity: 94.59%      Conservative: 34
Best Local Similarity: 89.88%      Mismatches: 31
Query Match:      75.28%      Indels:      8
DB:              4              Gaps:      4

US-09-610-313B-32 (1-2457) x US-09-309-572-17 (1-1003)

```

```

QY      220  TTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGGCAAGGCCCGCGAGTTCCCGCAGCGAGCAG 279
Db      1   PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY      280  AACCGCGCCCAACAGCCCAACAGCCGCGAGCTGCAGGTG-----CGCGCGCAGCAACCCC 333
Db      21  ThrAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY      334  CGCAGCGAGCGCGCGCGCGAGCGCCAGGCGCACCCCTG-----AACTTCCCGCAGATCAC 387
Db      41  LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY      388  CTGTGGCAGCGCCCGCTGTGTGAGCATCAAGGTGGCGCGCAGATCAAGGAGGCCCTGTGCTG 447

```

```

Db      61   LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY      448  GACACCGCGCGCGAGCACACCGTGTGAGAGATGAGCCTGCCCGCAAGTGGAGAGCC 507
Db      81   AspThrGlyAlaAspAspThrValLeuGluMetAsnLeuProGlyArgTrpLysPro 100
QY      508  AAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGCGCCACAGTACGACCAGATCTCTGATC 567
Db      101  LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120
QY      568  GAGATCTCGCGCAGAGAGGCCATCGGCACCGTGTGTGATCGGCCCCACCCCGTGAACATC 627
Db      121  GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY      628  ATCGGCCCAACATCTGACCCAGCTGGGTGCACTTCCGAACTTCCCATCAGACCCCATC 687
Db      141  IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY      688  GAGACCGTGCCTGGAAGTGAAGCCCGCATGAGCGCCCAAGCTGAAGTGAAGCAGTGGGCC 747
Db      161  GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY      748  CTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGGAGAGATCGAGAGAGAGGCG 807
Db      181  LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY      808  AAGATCACCAAGATCGCGCCCGAGAACCCCTACACACCCCGTGTTCGCGCATCAAGAG 867
Db      201  LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY      868  AAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAAACAAGCGCACCCAG 927
Db      221  LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY      928  GACTTCTCGGAGGTGACGTGGGCATCCCGACCCCGCGGCTGAAGAAGAAGAGAGC 987
Db      241  AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260
QY      988  GTGACCGTGTGGAGCTGGCGGAGCCCTACTTTCAGCGTCCCGTGGAGAGGACTTCGCG 1047
Db      261  ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY      1048  AAGTACACCGCTTCCACCATCCCGACATCAACACAGACAGCCCGCGCATCCGCTACCCAG 1107
Db      281  LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY      1108  TACAACGTGTGCCCCAGGGCTGGAGGGCAGCCCGACATCTTCCAGAGCAGCATGACC 1167
Db      301  TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320
QY      1168  AAGATCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGTATCCAG-----GCC 1221
Db      321  LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY      1222  CCCCTGTACGTGGCGAGCAGCTGGAGATCGCGCAGACCGCGCCCAAGATCGAGGAGCTG 1281
Db      341  AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 360
QY      1282  CGCAAGCACCTGTCTGCTGGGTGTTACACCCCGCAAGAGACGACCAAGAGAGCC 1341
Db      361  ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY      1342  CCCTTCCCTG-----CCCATCGAGCTGCACCCCGCAAGTGCACCGTGCAGCCCATCGAG 1395
Db      381  ProPheLeuTrpMetGlyTyrGlnLeuHisProAspLysTrpThrValGlnProIleVal 400
QY      1396  CTGCGCCGAGAGAGAGCTGGACCGCTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAC 1455
Db      401  LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY      1456  TGGCGCAGCGCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGC 1515
Db      421  TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

```

```
QY 1516 GCCAAGGCCCTGACCGACATCGTCCCTGACCGAGGAGCGAGCTGGAAGCTGCGCCGAG 1575
Db 441 ThrLysAlaLeuThrGluValProLeuThrGluAlaGluLeuGluLeuAlaGlu 460
QY 1576 AACCGGAGATCCCGGAGCGCGTGCACCGCGTGACTACGACCCCGAGCAAGGACCTG 1635
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1636 CTGGCCGAGATCCAGAACGAGCGGCACACACAGTGGACCTACACAGATCTACACGAGGCC 1695
Db 481 IleAlaGluIleGlnLysGlnGlyGlnThrTrpThrTyrGlnIleTyrGlnGluPro 500
QY 1696 TTCAGAACCTGAGAACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGACGTG 1755
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520
QY 1756 AAGCAGCTGACCGAGCGCGTGCAGAGATCCCATGGAGCATCGTATCTGGGCGAG 1815
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540
QY 1816 ACCCCCAAGTTCGCTCCCATCCAGAACGAGACCTGGGAGACCTGGTGACCGACTAC 1875
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTyr 560
QY 1876 TGGCAGGCCACTGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTG 1935
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 580
QY 1936 TGCTACCACTGCGAGAGAGCCCATCATCCGCGCGGACCTTCTAGCTGGAGCGGCC 1995
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600
QY 1996 GCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTACCGACCGCGCGCGCGCAGAG 2055
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
QY 2056 ATCGTGACCTGACCGCAGACCAACCAACAGAACCGAGTGCAGGCGCATCCAGCTGCC 2115
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2116 CTGCAGACGCGCAGCGAGGTGATCATCTGACCGACGACGACGACGACGACGACGAC 2175
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2176 ATCCAGGCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2235
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2236 ATCAAGAGGAGAGAGGTGTACTGAGCTGGTGGTCCCGCCCAACAGGGCATCGCGGCAAC 2295
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2296 GAGCAGATCGACAGCTGCTGAGCAGGAGGATCGCAGAGTGTCTCTGAGCGGATC 2355
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2356 GAT 2358
Db 721 Asp 721
```

## RESULT 8

```
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718.096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
```

```
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-718-096-17
```

```
Alignment Scores:
Pred. No.: 3,66-206 Length: 1003
Score: 3429,00 Matches: 648
Percent Similarity: 94.59% Conservative: 34
Best Local Similarity: 89.88% Mismatches: 31
Query Match: 75.28% Indels: 8
DB: 4 Gaps: 4
```

US-09-610-313B-32 (1-2457) x US-09-718-096-17 (1-1003)

```
QY 220 TTCCTCCGCGAGGACCTGGCTTCCCGCAGGCAAGGCCCGAGTTCCCGCAGGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCACACGCCCCACCGCCGCGAGCTGCAGGTG-----CGCGGCGACACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSer 40
QY 334 CGCAGCGAGGCGCGCGCGCGCGAGCGCGCGACCGCTG-----AACTTCCCGCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCTCGGTGAGCATCAAGTGGCGGCCGACATCAAGAGCGCCCTGTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCGCGCGCGCTGCGAGGAGTACGCTGCCCGGCAAGTGGNAGCCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluLeuMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCGCATGACGACCATCTGTATC 567
Db 101 LysMetIleGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeu 120
QY 568 GAGATCTGCGGCAAGAGGCCATCGGCACCGCTGTGATCGCGCCCGCCCGCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATCGTGTGACCCAGCTGGTGGTGCACCTGAACTTCCCATCAGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCTGCGCGGAGGAGTGAAGCCCGCATGACGCGCCCGCCCGAGGTGAAGCAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCCCATCTCGGAGGAGATGGAGAGGAGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGCGAGAACCCCTACACACACCCCGCTGTCGCCCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCAGCAGTGGCGCAAGCTGTGTGACTTCGCGAGCTGAACAAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGGAGGTGCAGCTGGGCGCATCCCCCGCGCGCTGAAAGAGAAGAGAGC 987
```





US-07-743-357-5

**Alignment Scores:**

Pred. No.:	5,56e-206	Length:	1016
Score:	3426.00	Matches:	648
Percent Similarity:	93.3%	Conservative:	36
Best Local Similarity:	88.4%	Mismatches:	29
Query Match:	75.21%	Indels:	20
DB:	2	Gaps:	5

US-09-610-313B-32 (1-2457) x US-07-743-357-5 (1-1016)

Qy	220	TTCTTCCGCGAGGACCTGGCCCTTCCCCAGGCGCAAGCCCGCAGTTC-----	267
Db	1	PhePheArgGluAAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
Qy	268	-----CCGACGCGAGCAGAACCGGCCCAACGAGCCCCCACCAGC	303
Db	21	ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg	40
Qy	304	CGCGAGCTGCAGGTG-----CGCGGCACAAACCCCGCAGCGAGCGCGGCCCGCAGCGC	357
Db	41	ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArg	60
Qy	358	CAGGCGACCCCTG-----AACTTCCCCAGATCACCTGTGTGGCAGCGCCCCCTGGTGAGC	411
Db	61	GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr	80
Qy	412	ATCAAGGTGGCGGCCGACATCAAGGAGGCCCTGTCTGGACACCGCGGCCGACGACACCGTG	471
Db	81	IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal	100
Qy	472	CTGCGAGGAGATGAGCTCCCGGCAAGTGGGAAGCCCAAGATGATCGGGGGCATCGCGCGC	531
Db	101	LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly	120
Qy	532	TTCATCAAGGTGCGCCAGTACGACACGAGATCTCTGATCGAGATCTGCGGCAAGAAGCCCATC	591
Db	121	PheIleLysValArgGlnTyAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle	140
Qy	592	GGCACCGTGTGATCGGCCCCACCCCGTGAACATCATCTCGCGCGCAACATGCTGACCCAG	651
Db	141	GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln	160
Qy	652	CTGGCGTGCACCTGAACCTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAG	711
Db	161	IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys	180
Qy	712	CCCGGCATGGACGCGCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAATCAAGGCC	771
Db	181	ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluLysIleLysAla	200
Qy	772	CTGACCGCCATCTCGGAGGAGATGGAGAAGGAGGCAAGATCAACAAGATCGGCCCGCAG	831
Db	201	LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu	220
Qy	832	AAACCCCTACACACCCCGTGTTCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCGCAAG	891
Db	221	AsnProTyAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys	240
Qy	892	CTGTGTGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTTGGAGGTGCAGCTGGGC	951
Db	241	LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly	260
Qy	952	ATCCCCACCCCGCGGCTGAAGAAGAAGAGCGTGACCGTCTGGACCTGGCGCGAC	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp	280
Qy	1012	GCCTACTTTCAGCGTCCCTCGACGAGGACTTCGCGAAGTACACCGCTTCACCATCCCC	1071
Db	281	AlaTyThrPheSerValProLeuAspGluAspPheArgLysTyThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACACGAGACCCCGGATCCGGCTACCAGTACAACGTGCTGCTCCCGAGGGCTGG	1131

Db	301	Ser	Ile	Asn	Asn	Glu	Thr	Pro	Gly	Ile	Arg	Tyr	Gln	Tyr	Asn	Val	Leu	Pro	Gln	Gly	Trp	320	
QY	1132	AAG	G	G	C	A	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1191	
Db	321	Lys	Gly	Ser	Pro	Ala	Ile	Phe	Gln	Ser	Met	Thr	Lys	Ile	Leu	Glu	Pro	Phe	Arg	Lys	340		
QY	1192	CG	C	A	C	A	C	C	C	C	G	A	T	C	G	T	A	C	C	C	C	1245	
Db	341	Gln	Asn	Pro	Asp	Ile	Val	Ile	Tyr	Gln	Tyr	Met	Asp	Asp	Leu	Tyr	Val	Gly	Ser	Asp	Leu	360	
QY	1246	GAG	A	T	C	G	G	C	A	G	C	T	G	C	A	G	A	T	C	G	C	1305	
Db	361	Glu	Ile	Gly	Gln	His	Arg	Thr	Lys	Ile	Glu	Glu	Leu	Arg	Gln	His	Leu	Leu	Arg	Trp	Gly	380	
QY	1306	TT	C	A	C	A	C	C	C	C	C	C	A	G	A	G	A	G	C	C	C	1359	
Db	381	Leu	Thr	Thr	Pro	Asp	Ile	Val	Lys	His	Gln	Lys	Glu	Pro	Pro	Phe	Leu	Trp	Met	Gly	Tyr	Glu	400
QY	1360	CT	G	A	C	C	C	C	C	A	A	G	T	G	C	A	C	C	C	C	C	1419	
Db	401	Leu	His	Pro	Asp	Ile	Val	Thr	Val	Gln	Pro	Ile	Val	Leu	Pro	Glu	Lys	Asp	Ser	Trp	Thr	420	
QY	1420	GT	G	A	A	C	A	T	C	C	A	A	G	T	G	T	G	G	C	A	A	1479	
Db	421	Val	Asn	Asp	Ile	Gln	Lys	Leu	Val	Gly	Lys	Leu	Asn	Trp	Ala	Ser	Gln	Ile	Tyr	Pro	Gly	440	
QY	1480	AT	C	A	G	T	G	C	G	C	A	G	T	C	T	G	C	G	G	C	C	1539	
Db	441	Ile	Lys	Val	Arg	Gln	Leu	Cys	Lys	Leu	Leu	Arg	Gly	Thr	Lys	Ala	Leu	Thr	Glu	Val	Ile	460	
QY	1540	CCC	T	G	A	C	C	A	G	A	G	G	C	C	G	A	G	A	C	C	G	1599	
Db	461	Pro	Leu	Thr	Glu	Glu	Ala	Glu	Leu	Glu	Leu	Ala	Glu	Asn	Arg	Ile	Ile	Leu	Lys	Glu	Pro	480	
QY	1600	GT	G	A	C	G	C	T	G	T	A	C	A	G	A	C	C	C	A	G	A	1659	
Db	481	Val	His	Gly	Val	Tyr	Tyr	Asp	Pro	Ser	Lys	Asp	Leu	Ile	Ala	Glu	Ile	Gln	Lys	Gln	Gly	500	
QY	1660	CAC	G	A	C	A	G	T	G	A	C	T	A	C	C	A	G	A	C	C	C	1719	
Db	501	Gln	Gly	Gln	Trp	Thr	Tyr	Gln	Ile	Tyr	Gln	Glu	Pro	Phe	Lys	Asn	Leu	Lys	Thr	Gly	Lys	520	
QY	1720	TAG	C	C	A	A	G	A	T	G	C	C	A	C	C	A	C	A	C	A	G	1779	
Db	521	Tyr	Ala	Arg	Thr	Arg	Gly	Ala	His	Thr	Asn	Asp	Val	Lys	Gln	Leu	Thr	Glu	Ala	Val	Gln	540	
QY	1780	AAG	A	T	C	C	C	A	T	C	G	T	G	G	C	A	A	C	C	C	C	1839	
Db	541	Lys	Ile	Thr	Thr	Glu	Ser	Ile	Val	Ile	Trp	Gly	Lys	Thr	Pro	Lys	Phe	Lys	Leu	Pro	Ile	560	
QY	1840	CAG	A	G	A	G	A	C	T	G	G	A	C	T	G	T	G	C	A	C	C	1899	
Db	561	Gln	Lys	Glu	Thr	Trp	Glu	Thr	Trp	Thr	Glu	Tyr	Trp	Gln	Ala	Thr	Trp	Ile	Pro	Glu	580		
QY	1900	TGG	G	A	G	T	T	C	T	G	A	A	C	C	C	C	C	C	C	C	C	1959	
Db	581	Trp	Glu	Phe	Val	Asn	Thr	Pro	Pro	Leu	Val	Lys	Leu	Trp	Tyr	Gln</							

Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
 QY 2200 GAGAGCGAGCTGGTGAACAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACCTG 2259  
 Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGluLysValTyrLeu 700  
 QY 2260 ACTCGGTGCGCGCCACAGGCGCATCGCGGCAAGCAGCAGCATCGACAAGCTGTGAGC 2319  
 Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
 QY 2320 AAGGGCATCCGCAAGGTGCTGTCTCTGACGCGCATCGAT 2358  
 Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 10  
 US-07-743-357-3  
 ; Sequence 3, Application US/07743357  
 ; Patent No. 5858646  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kang, Yong C.  
 ; TITLE OF INVENTION: Polypeptide having immunological  
 ; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: KIRBY EADES GALE BAKER  
 ; STREET: Box 3432, Station D  
 ; CITY: Ottawa  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: K1M 1H8  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/743,357  
 ; FILING DATE: 21-AUG-1991  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/CA90/00062  
 ; FILING DATE: 23-FEB-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gale, Edwin J.  
 ; REGISTRATION NUMBER: 28,584  
 ; REFERENCE/DOCKET NUMBER: 30924-2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (613) 237-6900  
 ; TELEFAX: (613) 237-0045  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1016 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; FRAGMENT TYPE: Internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Human immunodeficiency virus type 1  
 ; STRAIN: BH5  
 US-07-743-357-3

Alignment Scores:  
 Pred. No.: 6,41e-206 Length: 1016  
 Score: 3425.00 Matches: 648  
 Percent Similarity: 93.18% Conservative: 35  
 Best Local Similarity: 88.40% Mismatches: 30  
 Query Match: 75.19% Indels: 20  
 DB: 2 Gaps: 5  
 US-09-610-313b-32 (1-2457) x US-07-743-357-3 (1-1016)

QY 220 TTCTTCCGCGAGGACTCGGCTTCCCCAGAGGCGCCGCGAGTTC----- 267  
 Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
 QY 268 -----CCACAGCAGCAGAACCCGCGCCCAACAGCCCCACAGC 303  
 Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
 QY 304 CGCAGCTGCGAGTG-----CGCGGCACAAACCCCGCAGCAGCGCGCGCGCGCAGCGC 357  
 Db 41 ArgGluLeuGlnValTyrGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
 QY 358 CAGGSCACCTCG-----AACTTCCCCAGATCACCTGTGCGACGCGCCCTGTGTGAGC 411  
 Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
 QY 412 ATCAAGTGGCGCGCAGATCAAGAGGCGCTGTGTGGACACCGCGCGCGCAGACACCGTG 471  
 Db 81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
 QY 472 CTGGAGGAGATGAGCTGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531  
 Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120  
 QY 532 TTCTATCAAGGTGCGCAGCTACGACGATCTGTATCGAGATCTGCGGCAAGAGCCATC 591  
 Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
 QY 592 GGCACCGTGTGATCGCGCCCAACCGCTGAACATCATCGCGCGCAACATGCTGACCCAG 651  
 Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
 QY 652 CTGGCTGCACCTCGAACTTCCCATCAGCCCATCGAGACGCTGCCGTGAAGTGAAG 711  
 Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
 QY 712 CCCGCGATGGAGCGGCCCAAGGTGAAGTGAAGTGGCCCTGACCGAGAGAGATCAAGGCC 771  
 Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
 QY 772 CTGACCGCGCATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCGCAG 831  
 Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
 QY 832 AACCCCTACAACACCCCGCTGTTCGCCATCAAGAGAGAGCAGCAGCAAGTGGCGCAG 891  
 Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
 QY 892 CTGTTGGACTTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGTGCAGCTGGGC 951  
 Db 241 LeuValAspPheArgGluLeuAsnArgThrGlnAspPheTrpGluValGlnLeuGly 260  
 QY 952 ATCCCCCAACCGCGCGCTGAAGAGAGAGCGTGCACCGTCTGACCTGGCGCGAC 1011  
 Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
 QY 1012 GCCTACTTACGCGTCCCTGAGCAGGACTTCCGCAAGTACACCGCTTACACCTCCCGC 1071  
 Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
 QY 1072 AGCATCAACACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCGCCAGCGCTGG 1131  
 Db 301 SerIleAsnAsnGluThrProGlySerGlyTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
 QY 1132 AAGGCGACCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGCCCTTCGCGCGC 1191  
 Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340  
 QY 1192 CGCAACCGCGAGATCGTGTATCTACCG-----GCCCGCTGTAGCTGGCGAGCGACCTG 1245  
 Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
 QY 1246 GAGATCGGCGCAGCACCGCGCCCAAGATCGAGAGCTGCGCAAGACCTGTGCGCTGGGCGC 1305





ADDRESSEE: KIRBY EADES GALE BAKER  
 STREET: Box 3432, Station D  
 CITY: Ottawa  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: K1M 1H8  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/743,357  
 FILING DATE: 21-AUG-1991  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/CA90/00062  
 FILING DATE: 23-FEB-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gale, Edwin J.  
 REGISTRATION NUMBER: 28,584  
 REFERENCE/DOCKET NUMBER: 30924-2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (613) 237-6900  
 TELEFAX: (613) 237-0045  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1003 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 FRAGMENT TYPE: internal  
 ORIGINAL SOURCE:  
 ORGANISM: Human immunodeficiency virus type 1  
 STRAIN: ELI  
 US-07-743-357-10

## Alignment Scores:

Pred. No.: 2,17e-205 Length: 1003  
 Score: 3416.50 Matches: 642  
 Percent Similarity: 94.72% Conservative: 40  
 Best Local Similarity: 89.17% Mismatches: 31  
 Query Match: 75.01% Indels: 7  
 DB: 2 Gaps: 4

US-09-610-313b-32 (1-2457) x US-07-743-357-10 (1-1003)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCTCCAGGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279  
 DB 1 PhePheArgGluAenLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerProLysGln 20  
 QY 280 AACCGCGCCACAGCCGCCACCGAGCGGAGCTGCGCGC-----GACACCCCGCG 336  
 DB 21 ThrArgAlaAenSerProThrSerArgGluLeuArgValTrpGlyArgAspAenProLeu 40  
 QY 337 AGCGAGGCGCGCGCGAGCGCGGCGGACCCCTG-----AACTTCCCGCAGATCACCCCTG 390  
 DB 41 SerLysThrGlyAlaGluArgGlnGlyThrValSerPheAenPheProGlnIleThrLeu 60  
 QY 391 TGGCAGCGCCCTGGTGAGCATCAAGTGGCGGCGGCGAGTCAAGAGGCGCCCTGCTGGAC 450  
 DB 61 TrpGlnArgProLeuValAlaIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 80  
 QY 451 ACCGCGCGCCACGACCCGCTGGAGAGATGAGCTGCCCGCAGTGGAGCCCAAG 510  
 DB 81 ThrGlyAlaAspThrValLeuGluGluMetAenLeuProGlyLysTrpLysProLys 100  
 QY 511 ATGATCGGCGGCTCGCGGCTTCATCAAGTGGCGGCGGAGTACGACACGATCCTGATCGAG 570  
 DB 101 MetIleGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleProIleGlu 120

QY 571 ATCTGCGGCAAGAGGCGCATCGGCACCGTGCTGTGATCGGCCCCACCCCGTGAACATCATC 630  
 DB 121 IleCySGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAenIleIle 140  
 QY 631 GGCGCAACATGCTGACCCAGCTGGCGGTGGACCTGAACCTCCCATCAGCCCATCAG 690  
 DB 141 GlyArgAenLeuLeuThrGlnIleGlyCysThrLeuAenPheProIleSerProIleGlu 160  
 QY 691 ACCGTGCCCTGAAAGCTGAAGCCCGCATGACCGCGCCCAAGAGTGAAGAGTGGCCCTG 750  
 DB 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180  
 QY 751 ACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGGAGGAGATGGAAGAGAGGCGCAAG 810  
 DB 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys 200  
 QY 811 ATCACCAGATCGGCGCGGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAGAG 870  
 DB 201 IleSerArgIleGlyProGluAenProIlePheAenIleLysLysLys 220  
 QY 871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGAGCTGAACAAGCGCACCCAGGAC 930  
 DB 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAenLysArgThrGlnAsp 240  
 QY 931 TTCTGGAGGTGACGTGGGCATCCCCACCCCGCGCTGAAGAAGAGAGAGCGGTG 990  
 DB 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
 QY 991 ACCGTGTGACGTGGCGGAGCTTACTTCAAGCGTGGCGCTGGAGGAGTCCCGCAAG 1050  
 DB 261 ThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAenPheArgLys 280  
 QY 1051 TACACCGCTTCCACCATCCCGCATCAACAAGAGACCCCGCATCCCGTACCATGATC 1110  
 DB 281 TyrThrAlaPheThrIleSerSerIleAenAenGluThrProGlyIleArgTrpGlnTyr 300  
 QY 1111 AAGTGTGCTGCCAGGCTGGAAGGCGAGCGCCAGCATCTTCCAGAGCAGCATGACCAAG 1170  
 DB 301 AenValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 QY 1171 ATCTGGAGCGCTTCCGCGCGCGCATCCCGAGATCGTGATCTACCAAG-----GCCCCC 1224  
 DB 321 IleLeuGluProPheArgLysGlnAenProGluMetValIleTrpGlnTrpMetAspAsp 340  
 QY 1225 CTGTACGTGGCGCAGCACCTGGAGATCGGCGAGCACCGCGCGCAAGATCGAGAGCTGCGC 1284  
 DB 341 LeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg 360  
 QY 1285 AAGCAGCTGTGCGTGGCGCTTCCACACCCCGCAAGAAGACCAAGAGGAGCGCCCGCC 1344  
 DB 361 GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro 380  
 QY 1345 TTCTGT-----CCCATCGAGCTGACCCCGAGAGTGGACCGTGGCGCCCATCGAGCTG 1398  
 DB 381 PheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400  
 QY 1399 CCGGAGAGAGAGAGCTGGACCGGTGAACACATCCAGAAAGCTGGTGGCGCAAGCTCAACTGG 1458  
 DB 401 ProGluLysGluSerTrpThrValAenAspIleGlnAenLeuValGluArgLeuAenTrp 420  
 QY 1459 GCCAGCGAGATCTACCCCGGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGCGCGCC 1518  
 DB 421 AlaSerGlnIleTrpProGlyIleLysValArgGlnLeuLysLysLeuLeuArgGlyThr 440  
 QY 1519 AAGCGCTGACACACATCGTGGCGCGCGAGGAGCGGAGTGGAGCTGGCGCGAGAAC 1578  
 DB 441 LysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGluAen 460  
 QY 1579 CGGAGATCTCGCGAGCGCGTGGCGGCTGTACTACGACCCCGACAGAGCTGGTGGTGG 1638  
 DB 461 ArgGluLysLysGluProValHisGlyValTrpTrpAspProSerLysAspLeuIle 480  
 QY 1639 GCCGAGATCCAGAAAGCGAGGCGCGACCGAGTGGACCTTACAGATCTTACAGGAGCCCTTC 1698



481	Db	 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe	500
1699	QY	AAGAACTCTGAAGACCGGCAAGTAGTCGCAAGATGCGCACCGCCACACCAACGACGCTGAAG	1758
501	Db	 LysAsnLeuLysThrGlyLysTyrAlaAspWeArgGlyAlaHisThrAsnAspValLys	520
1759	QY	CAGCTGACGAGCGCGTCGCAAGATCGCATGGAGAGCATCGTGTACTCTGGGCGCAAGACC	1818
521	Db	 GlnLeuAlaGluAlaValGlnArgIleSerThrGluSerIleValIleTrpGlyArgThr	540
1819	QY	CCCAAGTTCGCTCGCCATCCAGAGGAGACCTGGGAGACCTGGTGGAACCGACTACTGG	1878
541	Db	 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpAlaGluTyrTrp	560
1879	QY	CAGGCCACTCTGGATCCCGAGTGGGAGTCTCGTGAACACCCGCCCTGGTGAAGCTGTGG	1938
561	Db	 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp	580
1939	QY	TACCAGCTGAGAGAGGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGGCC	1998
581	Db	 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla	600
1999	QY	AACCGCGAGACCAAGATCGCGAAGGCGCGCTACGTGACCCACCGGGGCGCGGAGAAGATC	2058
601	Db	 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal	620
2059	QY	GTGAGCTGTACCGAGACACCAACAGACCGAGCTCGAGGCGCATCCAGCTGGCCCTG	2118
621	Db	 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu	640
2119	QY	CAGGACAGCGGACGAGGTGAACATCGTGACCGACAGCCAGTACGCGCTGGGCGATCATC	2178
641	Db	 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle	660
2179	QY	CAGGCCAGCGCCACAGACGAGCGAGCTGGTGAACCATCATCATCGAGCAGCTGATC	2238
661	Db	 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIle	680
2239	QY	AAGAAGGAGAAAGGTGTACTCTGAGCTGGGTGCGCGCCACCAAGGCGCATCGCGCGCAACGAG	2298
681	Db	 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu	700
2299	QY	CAGATCGCAAGCTGTGTAGACAAGGCGATCCGCAAGGTGTGTCTCTGGACGGCATCAT	2358
701	Db	 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp	720

RESULT 13

```

US-07-743-357-8
; Sequence 8, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

```

QY 811 ATCCAGAGTGGCGCCCGAGAACCCCTTACCAACACCCCGTGTTCGCATCAAGAGAG 870  
DB 201 ILeSerLySileGlyProGluAAsPProTyrAsnThrProValPheAlaIleLySLeYs 220  
QY 871 GACAGCACCAAGTCGCGCAAGCTGGTACCTTCGCGAGCTGAACAGCGCACCCAGGAC 930  
DB 221 AspSerThrLySAsPProGlySLeuValAsPProGlySLeuAAsPProGlySLeuAAsP 240  
QY 931 TTCTGGAGGTGTCAGCTGGGATCCCTCCACCCCGCGCTTGAAGAAGAGAGCGGTG 990  
DB 241 PheTrpGluValGlnLeuGlyLeProHisProAlaGlyLeuLySLeuLySLeuVal 260  
QY 991 ACCGTGTGACGTGGGAGCGCTTACCTTACGCTGCGCTGGAGCAGGACTTCGCGAAG 1050  
DB 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLySLeuPheArgLyS 280  
QY 1051 TACACCGCTTACCATCCCGAGCATCAACCAACAGACCCCGCATCCGCTACCATAC 1110  
DB 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProArgIleArgTyrGlnTyr 300  
QY 1111 AACGTGTGCTGCCCGAGGCTGAAGGCGAGCCCGACCATCTTCCAGAGCAGCATGACCAAG 1170  
DB 301 AsnValLeuProGlnGlyTrpLySLeuGlySerProAlaIlePheGlnSerSerMetThrLyS 320  
QY 1171 ATCTGTGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATACAG-----GCCGCC 1224  
DB 321 ILeuGluProPheLySLeuGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
QY 1225 CTGTACGTGGCAGCGGACCTGGAGATCGGCGAGCAGCCGCGCAAGATCGAGGACTGGC 1284  
DB 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgIleLySLeuGluLeuArg 360  
QY 1285 AAGCACTGTCTGCTGGGCTTCCACACCCCGCAAGAGACCAAGAGGAGCCCGCC 1344  
DB 361 GluHisLeuLeuLySLeuTrpGlyPheThrThrProAspLySLeuHisGlnLySLeuProPro 380  
QY 1345 TTCTCTG-----CCATCGAGCTGACCCCGCAAGTGGACCGTGGAGCTGCGAGCTG 1398  
DB 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLySLeuThrValGlnProIleValLeu 400  
QY 1399 CCGAGAGAGGAGCTGGACCGTGAACGACATCCAGATCGAGCTGGTGGCGAGCTGAATGG 1458  
DB 401 ProGluLySAspSerTrpThrValAsnAspIleGlnLySLeuValGlyLySLeuAsnTrp 420  
QY 1459 GCCAGCAGATCTACCCCGGATCAAGTGGCGGCGAGCTGTCAAGCTGCTGCGGGCGCC 1518  
DB 421 AlaSerGlnIleTyrAlaGlyIleLySValySLeuGlnLeuCySLeuLeuArgLyThr 440  
QY 1519 AAGCCCTGACCGACATCTGTCCTGACCGAGAGGCGCGAGCTGGAGCTGGCGAGAAC 1578  
DB 441 LySAlaLeuThrGluValValGlnLeuThrLySLeuAlaGluLeuAlaGluAsn 460  
QY 1579 GCGAGATCTGCGCGAGCGCGTGCAGCGGTGTACTACGACCCCGAGAGGAGCTGGTG 1638  
DB 461 ArgGluIleLeuLySLeuProValHisGlyValTyrTyrAspProSerLySAspLeuIle 480  
QY 1639 GCCAGATCCAGAGAGGCGCACGACGAGTGGACCTTACAGATCTACGAGGCGCTTC 1698  
DB 481 AlaGluIleGlnLySLeuGlyGlnTyrTrpThrTyrGlnIleTyrGlnGluProPhe 500  
QY 1699 AAGAACTGAAGACCGGCAAGTACGCAAGATGGCGACCGCCACACCAACGAGCTGAAG 1758  
DB 501 LySAsnLeuLySLeuThrGlyLySLeuTyrAlaArgMetArgGlyAlaHisThrAsnAspValLyS 520  
QY 1759 CAGTGTACCGAGGCGGTGCGAGAAGATCGCCATGAGAGAGCATCGTGTATCTGGGCGAAGCC 1818  
DB 521 GlnLeuThrGluAlaValGlnLySValAlaThrGluSerIleValIleTrpGlyLySLeu 540  
QY 1819 CCCAAGTTCGCTGCGCTTCCATCCAGAGGAGACCTGGGAGACCTGGTGGACGACTACTGG 1878  
DB 541 ProLySLeuProIleGlnLySLeuThrTrpGluAlaTrpTrpThrGluTyrTrp 560  
QY 1879 CAGGCCACCTGGATCCCGCGAGTGGAGTTCGTGAACACACCCCGCTGGTGAAGCTGTGG 1938

DB 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLySLeuTrp 580  
QY 1939 TACCAGCTGGAGAGGAGCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGGCGCCGCC 1998  
DB 581 TyrGlnLeuGluLySLeuGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 1999 AACCGCGAGACCAAGATCGGCAAGCGCGGTGTACGTGACCGACCGGGCGCGCAGAAGATC 2058  
DB 601 AsnArgGluThrLySLeuGlyLySLeuAlaGlyTyrValThrAspArgGlyArgGlnLySVal 620  
QY 2059 GTGAGCCTGACCGAGAGACCAACCAAGACCGAGCTGCGAGGCTGCGAGCTGCGCCTG 2118  
DB 621 ValSerLeuThrAspThrThrAsnGlnLySLeuGlnAlaIleHisLeuAlaLeu 640  
QY 2119 CAGGACAGCGCGCAGCAGGTGAAACATCGTACCGACGACGAGTACGCGCTGGGCATCATC 2178  
DB 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
QY 2179 CAGCCCGAGCCCGCAAGAGCGAGCGAGCTGGTGAACACGAGATCATCGAGCAGCTGATC 2238  
DB 661 GlnAlaGlnProAspLySLeuGluValSerGlnIleIleGluGlnLeuIle 680  
QY 2239 AAGAGAGAGAGGTGTACTGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2298  
DB 681 LySLeuGluLySLeuValTyrLeuAlaTrpValProAlaHisLySLeuGlyGlyAsnGlu 700  
QY 2299 CAGATCCACAGCTGGTGGAGCAGGCGATCGCAAGGTGTGTCTTCTGGAGCGGATCGAT 2358  
DB 701 GlnValAspArgLeuValSerThrGlyIleArgLySValLeuPheLeuAspGlyIleAsp 720

RESULT 14  
US-07-743-357-7  
; Sequence 7, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1004 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein

; HYPOTHETICAL: NO  
 ; FRAGMENT TYPE: Internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Human immunodeficiency virus type 1  
 ; STRAIN: SF2  
 ; US-07-743-357-7

Alignment Scores:  
 Pred. No.: 9, 82e-205 Length: 1004  
 Score: 3406.00 Matches: 642  
 Percent Similarity: 94.31% Conservative: 38  
 Best Local Similarity: 89.04% Mismatches: 33  
 Query Match: 74.77% Indels: 8  
 DB: 2 Gaps: 4

US-09-610-313B-32 (1-2457) x US-07-743-357-7 (1-1004)

QY	220	TTCTTCCGAGGAGCTGCTGCTCCCGCAGGCAAGGCGCGAGTCTCCCGAGGAGCAG	279
DB	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
QY	280	AACCGGCGCAACAGCCCAACAGCGCGAGCTGCAGTGCAGTGCAGGCGCGC-----GACAAACCC	333
DB	21	ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer	40
QY	334	CGCAGCGAGGCGCGCGAGCGCGAGCGCGAGCGCGCGCTG-----AAGTCTCCCGCAGATCACC	387
DB	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnProGlnIleThr	60
QY	388	CTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGCGCGCGAGATCAAGGAGGCGCTGCTG	447
DB	61	LeuTrpGlnArgProLeuValThrIleArgIleGlyGlnLeuLysGluAlaLeuLeu	80
QY	448	GACACCGGCGCGAGCAGACACCTGTGTGAGGAGATAGCTCCCGCGCAAGTGGAGCGCC	507
DB	81	AspThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro	100
QY	508	AAAGATGATCGGGCGCATCGCGCGCTTCATCAAGTGGCGCGAGTACGACAGATCTGTATC	567
DB	101	LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleProVal	120
QY	568	GAGATCTGCGCAAGAAGCCATCGGCACCGTGTGATCGCGCGCCCGCCCGTGAACATC	627
DB	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
QY	628	ATCGCGCGCAACATGCTGACCGAGCTGGGCTGCACCTGAACTTCCCATCGACCCCATC	687
DB	141	IleGlyArgAsnLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
QY	688	GAGACCGTCCGTGAAGCTGAAGCGCGCATGGACCGGCCCAAGTGAAGCAGTGGCGCC	747
DB	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
QY	748	CTGACCGAGGAGAAGATCAAGCGCTCTGACCGGCATCTGCGAGGAGATGAGAGGAGGCG	807
DB	181	LeuThrGluLysLysLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
QY	808	AAAGATCACCAAGATCGGCGCGCGAGAACCCCTCAACACCCCGTGTGTGCGCATCAAGAG	867
DB	201	LysIleSerLysIleGlyProGluAsnProThrAsnThrProValPheAlaIleLysLys	220
QY	868	AAGGACAGCACAAGTGGCGCAAGCTGGTGGACTTCCCGAGCTGAACAAGCGCACCCAG	927
DB	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
QY	928	GACTTCTGGAGGTGAGCTGGGCGATCCCGCCCGCGCGCTGAAGAAGAAGAAGAGC	987
DB	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
QY	988	GTCAGCTGCTGGAGTGGGCGAGCGCTTACTTTCAGGTGCCCTGGACGAGGACTTCCGC	1047
DB	261	ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheArg	280

QY	1048	AAGTACACCGCTTCCACCATCCCGAGCATCAACACGAGACCCCGCGCATCCGCTACCAG	1107
DB	281	LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyLeuArgTyGln	300
QY	1108	TACAACTGCTGCGCGCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC	1167
DB	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320
QY	1168	AAGATCTCTGAGCGCTTCCCGCGCGCAACCCCGAGATCGTGATCTACCAG-----GCC	1221
DB	321	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyGlnTrpMetAsp	340
QY	1222	CCCTGTGACGTGGCGAGCGACCTGAGATCGCGCAGCACCGCGCAAGATCGAGGAGCTG	1281
DB	341	AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
QY	1282	CGAAGACCTGTGCTGGCTGGGGCTTACACACCCCGCAAGAGACACAGAGGAGGCC	1341
DB	361	ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
QY	1342	CCCTTCTCTG-----CCCATCGAGCTGCACCCCGACAAAGTGCAGCGTCCCATCGAG	1395
DB	381	ProPheLeuTrpMetGlyTyThrGluLeuHisProAspLysTrpThrValGlnProIleMet	400
QY	1396	CTGCCCGCAAGAGGAGAGCTGACCGTGAACGACATCCAGAAAGCTGGTGGCAAGCTGAAC	1455
DB	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
QY	1456	TGGGCGACCGAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGCAAGCTGCTGCGCGCG	1515
DB	421	TrpAlaSerGlnIleTyAlaGlyLysValLysGlnLeuCysLysLeuLeuArgGly	440
QY	1516	GCCAAAGCGCTGACGACATCGTCCCTGACCGAGGAGCGCGAGCTGCAGTGCGCCGAG	1575
DB	441	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu	460
QY	1576	AACCGCAGATCTTCCGCGAGCGCGTGCACCGCGCTGTACTACGACCCCGCAAGACCTG	1635
DB	461	AsnArgGluIleLeuLysGluProValHisGluValTyThrAspProSerLysAspLeu	480
QY	1636	GTGCGCGAGATCCAGAACGAGCGGCACACAGATGGAGCTTACCAGATCTTACCAGAGCC	1695
DB	481	ValAlaGluIleGlnLysGlnGlyGlnTyThrTyThrTyThrGlnIleTyThrGlnGluPro	500
QY	1696	TTCAGAACCTTGAAGACCGCGCAAGTACCGCAAGATGCGCACCGCCACACCAACGACGTG	1755
DB	501	PhenylAsnLeuLysThrGlyLysTyThrAlaArgMetArgGlyAlaHisThrAsnAspVal	520
QY	1756	AAGCAGCTGACCGAGCGCGTGCAGAAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG	1815
DB	521	LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys	540
QY	1816	ACCCCAAGTTCGCGCTGCCCATCCAGAAAGAGACCTGGGAGACCTGGTGGACCCACTAC	1875
DB	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTy	560
QY	1876	TGGCAGGCGACCTGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTG	1935
DB	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu	580
QY	1936	TGGTACCAAGCTGGAGAAGGAGCCCATCATCGCGCGCGAGACCTTCTAGCTGACGCGCGC	1995
DB	581	TrpTyThrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyThrValAspGlyAla	600
QY	1996	GCCAAACCGCAGAACAGATCGGCAAGCGCGGCTACGTGACCGACCGCGGCGCGCAGAG	2055
DB	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyThrValThrAspArgGlyArgGlnLys	620
QY	2056	ATCGTGCAGCTGACCGAGACCAACAGAGAGCCGAGCTGCAGGCGCATCCAGCTGGCC	2115
DB	621	ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla	640
QY	2116	CTGCAGGACAGCGCGCAGCGAGGTGAACATCGTGACCGACGACGACGTACCGCTGGGCATC	2175

```

Db      641 LeuGlnAAspSerGlyLeuGluValAsnIleValThrAAspSerGlnTyrAlaLeuGlyIle 660
Qy      2176 ATCCAGGCCCGAGCCGCAAGAGCGAGAGCGAGTGTGGTGAACAGATCATCGAGCAGCTG 2235
Db      661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
Qy      2236 ATCAAGAAGGAGAGGTGTACTGAGCTGGGTGGCCCGCCCAAGAGGATCGGCGGCAAC 2295
Db      681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
Qy      2296 GAGCAGATCGACAAGCTGGTCAGCAAGGGCATCCGCAAGGTGCTGTTCTCGAGCGGCATC 2355
Db      701 GluGlnValAAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720
Qy      2356 GAT 2358
Db      721 Asp 721

RESULT 15
US-09-319-588C-6
; Sequence 6, Application US/09319588C
; Patent No. 6509018
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAULIERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSI, Françoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/09/319,588C
; CURRENT FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-319-588C-6

Alignment Scores:
Pred. No.:      3,72e-188      Length:      1014
Score:          3140.50      Matches:      579
Percent Similarity: 89.70%      Conservative: 74
Best Local Similarity: 79.53%      Mismatches: 60
Query Match:    68.95%      Indels:      15
DB:             4           Gaps:         5

US-09-610-313B-32 (1-2457) x US-09-319-588C-6 (1-1014)
Qy      220 TTCTTCGCGAGGACCTGGCTTCCCGCCGAGGAGCCCGCGAGTTCCTCCCGAGCGAGCAG 279
Db      1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProProAspAsn 20
Qy      280 AAC-----CGCGCAACAGCCCAAGCGCGCGCGAGTGCAGGTG----- 318
Db      21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40
Qy      319 ---CGCGCGCAACCCCGCAGGAGCGCGCGCGCGCGCGCGCGAG-----GGCACC 366
Db      41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60
Qy      367 CTGAACCTCCCGCAGATCACCCTGTGGCAGCGCCCTGTGGTGGACATCAAGGTGGCGGC 426
Db      61 PheAsnPheProGlnIleThrLpnlArgProValIleThrValLysIleGlyLys 80
Qy      427 CAGATCAAGGAGGCGCCCTGCTGGACACCGCGCGCGAGCACACCGTGTGGAGGATGAGC 486

```

```

Db      81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluLeuGln 100
Qy      487 CTGCCCGCGCAAGTGCAGAGCCCAAGATCATCGCGCGCATCGCGGCTTCATCAAGGTGGCG 546
Db      101 LeuGluGlyLysTrpLysProLysMetIleGlyIleGlyLysPheIleLysValArg 120
Qy      547 CAGTACACACAGATCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGTATC 606
Db      121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140
Qy      607 GGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGGTGCACCCGTG 666
Db      141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160
Qy      667 AACTTCCCATCATCGACCGTGCAGACCGTCCCTGAAAGCTGAAGCCCGCGCATCGACGGC 726
Db      161 AsnPheProLysSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180
Qy      727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGC 786
Db      181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200
Qy      787 GAGGAGATGAGAAGGAGGCGCAAGATCACCAGATCGGCCCGCGAGAACCCCTACACACC 846
Db      201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220
Qy      847 CCCTGTTCCCATCAAGAAGAAGAGCAGCAGCAAGTGGCGCAAGCTGGTGGTACTCCGC 906
Db      221 ProfilePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240
Qy      907 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGCGCATCCCCACCCCGCC 966
Db      241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260
Qy      967 GGCTTGAAGAAGAAGAGAGCGTGACCGTGTGACGTGGCGCAGCGCTACTTCAGCGTG 1026
Db      261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280
Qy      1027 CCCTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAACAACGAG 1086
Db      281 ProLeuAspLysAspPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGlu 300
Qy      1087 ACCCGCGCATCGCTACCAAGTACCAAGTGTGCGCCAGAGGTGGAAGGGCAGCGCCCGAC 1146
Db      301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320
Qy      1147 ATCTTCAGAGCAGATGACCAAGATCCTGAGCGCTTCGCGCGCCGCGCAACCCCGAGATC 1206
Db      321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340
Qy      1207 GTGATCTACCAG-----GCCCGCTGTAGTGGCGCAGCGCTGGAGATCGGCCAGCAC 1260
Db      341 IleIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360
Qy      1261 CGCGCCCAAGATCGAGGAGCTGCGCAAGCATCTGTGCGCTGGGCTTCACACCCCGGAC 1320
Db      361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380
Qy      1321 AAGAGCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1374
Db      381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400
Qy      1375 TGACCGCTGAGCCATCGAGCTGCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434
Db      401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420
Qy      1435 AAGCTGTGGCGAAGCTGAACCTGGCGCAGCAGATCTACCCCGCATCAAGGTGGCGCAG 1494
Db      421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440
Qy      1495 CTGTGAAGTGTGTGGCGCGCAAGCGCTGACCGACATCGTGGCCCTGACCGAGGAG 1554

```

```

Db      441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460
QY      1555 GCCGAGCTGGAGCTGGCGAGAACCGCGAGATCCTGCGCGAGCCGCTGCGACGGCGTGTAC 1614
Db      461 AlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480
QY      1615 TACGACCCCGACGAAGCACTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAAGTGGACC 1674
Db      481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnTrpThr 500
QY      1675 TACCAATCTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGC 1734
Db      501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520
QY      1735 ACCGCCCCACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAAATGCCATGGAG 1794
Db      521 SerAlaHisThrAsnAspLysGlnLeuValGluValValArgLysValAlaThrGlu 540
QY      1795 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCCCGCTGCCCATCCAGAAGGAGACCTGG 1854
Db      541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560
QY      1855 GAGACTGTGTGACCGACTACTGGAGGCCACTGGATCCCGAGTCCGAGTGGAGTTCGTGAAC 1914
Db      561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580
QY      1915 ACCCCCCCTGGTGAAGCTGTGTACGAGCTGGAGAGGAGCCCATCATCGGCGCGGAG 1974
Db      581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600
QY      1975 ACCTTCTACGTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2034
Db      601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620
QY      2035 ACCGACCGGGCGCGCAGAGATCTGTGACCTGTGACCGAGACCAACCAAGAGACCGAG 2094
Db      621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640
QY      2095 CTGCAAGGCATCAGCTGGCCCTGCGAGACGCGGAGGAGTGAACATCGTGACCGAC 2154
Db      641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660
QY      2155 AGCCAGTACGCCCTGGCATCATCCAGGCCCGACCGCGACAGCGAGCGAGCGAGCTGGTG 2214
Db      661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680
QY      2215 AACCAAGATCATCAGAGCTGATCAAGAGAGAGAGGTGTACTGAGTGGGTGCCCGCC 2274
Db      681 SerGlnIleIleGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla 700
QY      2275 CACAGGGGCATCGGCGCAACGAGCAGATCGACAGCTGGTGCAGCAAGGGCATCCGCAAG 2334
Db      701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720
QY      2335 GTGCTGTCTCGACGCGCATCGAT 2358
Db      721 IleLeuPheLeuAspGlyIleGlu 728

```

Search completed: June 2, 2005, 05:13:50  
Job time : 106.917 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 03:13:23 ; Search time 34.0828 Seconds  
(without alignments)  
10815.332 Million cell updates/sec

Title: US-09-610-313B-30  
Perfect score: 4583  
Sequence: 1 gtcagccacacatggcgca.....gggttagcaccggtaattc 2469

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09610313/runat 31052005 155137 15173/app query.fasta\_1.7893  
-DB=Issued Patents AA -QPMT=faetan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CGN 1 1 101 @runat 31052005 155137 15173 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/iaa/6C.COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3535	77.1	1005	2	US-07-743-357-1
2	3531.5	77.1	1003	2	US-07-743-357-9
3	3513	76.7	1015	3	US-08-463-210-9
4	3513	76.7	1015	3	US-09-124-900-3
5	3513	76.7	1015	4	US-08-463-028-9
6	3513	76.7	1016	2	US-07-743-357-2
7	3510	76.6	1003	4	US-09-309-572-17
8	3510	76.6	1003	4	US-09-718-096-17
9	3507	76.5	1016	2	US-07-743-357-5
10	3506	76.5	1016	2	US-07-743-357-3
11	3503	76.4	1016	2	US-07-743-357-4
12	3497.5	76.3	1003	2	US-07-743-357-10

13	3494.5	76.2	1003	2	US-07-743-357-8
14	3487	76.1	1004	2	US-07-743-357-7
15	3221.5	70.3	1014	4	US-09-319-588C-6
16	3179	69.4	913	2	US-07-743-357-22
17	3135.5	68.4	1350	4	US-09-952-060-35
18	3116	68.0	913	2	US-07-743-357-6
19	2860.5	62.4	875	4	US-09-952-060-6
20	2855	62.3	850	4	US-09-952-060-2
21	2827	61.7	562	3	US-09-117-217-14
22	2827	61.7	562	4	US-09-735-487-14
23	2814.5	61.4	875	4	US-09-952-060-8
24	2809	61.3	850	4	US-09-952-060-4
25	2796	61.0	560	4	US-10-205-641-1
26	2545.5	55.5	1018	4	US-09-206-551-46
27	2521.5	55.0	1016	4	US-09-625-972-24
28	2270	49.5	441	4	US-09-991-258-16
29	2247	49.0	1055	2	US-08-659-251-5
30	2247	49.0	1055	3	US-09-256-490-5
31	2247	49.0	1055	5	PCT-US96-11445-5
32	2228.5	48.6	3080	6	5223423-4
33	2228.5	48.6	3080	6	5223423-4
34	2192	47.8	427	3	US-09-690-265-1
35	1593	34.8	327	3	US-08-679-493A-71
36	1590	34.7	314	1	US-08-583-446-6
37	1590	34.7	314	1	US-08-444-882-6
38	1590	34.7	314	2	US-08-389-459A-6
39	1590	34.7	314	3	US-08-987-867A-6
40	1576	34.4	327	3	US-08-679-493A-72
41	1349	29.4	1150	3	US-09-238-303-9
42	1349	29.4	1150	4	US-09-946-239-9
43	1335	29.1	261	6	5320958-3
44	1335	29.1	261	6	5320958-3
45	1320	28.8	1031	3	US-08-811-682-15

#### ALIGNMENTS

#### RESULT 1

US-07-743-357-1  
; Sequence 1, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 1:

Sequence 8, Appli	Sequence 7, Appli	Sequence 6, Appli	Sequence 5, Appli	Sequence 4, Appli	Sequence 3, Appli	Sequence 2, Appli	Sequence 1, Appli	Sequence 15, Appli
-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	--------------------



; SEQUENCE CHARACTERISTICS: ; LENGTH: 1005 amino acids ; TYPE: amino acid ; STRANDEDNESS: not relevant ; TOPOLOGY: not relevant ; MOLECULE TYPE: protein ; HYPOTHETICAL: NO ; FRAGMENT TYPE: internal ; ORIGINAL SOURCE: ; ORGANISM: Human immunodeficiency virus type 1 ; STRAIN: HXB2 ; US-07-743-357-1									
Alignment Scores:									
Pred. No.:	3	5e-213	Length:	1005					
Score:	3535.00	Matches:	659						
Percent Similarity:	96.12%	Conservative:	34						
Best Local Similarity:	91.40%	Mismatches:	24						
Query Match:	77.13%	Indels:	4						
DB:	2	Gaps:	2						
US-09-610-313B-30 (1-2469) x US-07-743-357-1 (1-1005)									
Qy	220	TTCTTCGCGAGGACCTGGCGCTTCCCGGCAAGGCCCGGAGTTCCCGACGAGCAG	279						
Db	2	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	21						
Qy	280	AACGGCGCCAAACGCCACCGCGGAGCGGAGCTGCAGTG-----CGCGCGCAACCCC	333						
Db	22	ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer	41						
Qy	334	CGAGCGAGCGCGCGCGAGCGCGAGCGGACCGACCTG-----AACTTCCCGCAGATCAC	387						
Db	42	ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr	61						
Qy	388	CTGTGGCAGCGCCCTCGTGAGCATCAAGTGGCGGCGCAGATCAAGAGGCGCTGCTG	447						
Db	62	LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu	81						
Qy	448	GACACCGCGCGCAGCAGCGCTGCGAGGATGAGCTGCCCGCAAGTGGAGGCC	507						
Db	82	AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro	101						
Qy	508	AAGATGATCGCGCGCATCGCGCTTCCATCAAGTGGCGGCGCAGTACGACGACGATCCTGATC	567						
Db	102	LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIle	121						
Qy	568	GAGATCTGCGGCAAGAGGCCATCGGCACCTGCTGATCGGCGCCCGCCCGTGAACATC	627						
Db	122	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	141						
Qy	628	ATCGCGCGCAACATCTGACCCAGCTGGCTGCACCTGACCTTCCCATCAGCCCCATC	687						
Db	142	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	161						
Qy	688	GAGACCTGCGCGTGAAGCTGAAGCCCGGATGAGCGCGCGCCCAAGGTGAAGCAGTGGCCC	747						
Db	162	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	181						
Qy	748	CTGACCGAGGAGAGATCAAGGCGCTGACCGCATCTCGAGGAGATGAGAGAGGCGC	807						
Db	182	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	201						
Qy	808	AAGATCACAGATCGCGCGCGAGAACCCCTACACACCCCGTGTTCGCGCATCAAGAAG	867						
Db	202	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	221						
Qy	868	AAGGACAGCAGCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGAACAAGCGCACCCAG	927						
Db	222	LysAspSerThrLysTrpAsgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	241						
Qy	928	GACTTCTGGAGGTGACGTGGGCATCCCCACCGCGCTCGAAGAGAGAGAGAGC	987						

Db	242	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	261
Qy	988	GTGACCGTGTGAGCTGGCGGACGCTACTTCCAGCGTGCCTCCGACGAGGACTTCCGC	1047
Db	262	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg	281
Qy	1048	AAGTACACCGCTTCCACCATCCCGCAGCATCAACAACAGAGACCCCGGCTACCGCTACCAG	1107
Db	282	LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln	301
Qy	1108	TACAACGTGTGCTGCCCGAGGCTGGAAGGAGGAGCCCGCAGCATCTTCCAGCAGCATGACC	1167
Db	302	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	321
Qy	1168	AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACAGTAGTACATGGAC	1227
Db	322	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp	341
Qy	1228	GACCTGTGATGGGCGAGCGACCTTGGAGATCGGCGCAGCACCGCGCCCAAGATCGAGGAGCTG	1287
Db	342	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	361
Qy	1288	CGCAAGACCTGTGCTGGCGCTTCCACACCCCGCAGCAAGAGCACCAGAGAGGCC	1347
Db	362	ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro	381
Qy	1348	CCCTTCTGTGGATGGCTACGAGCTGCACCCGACAAAGTGGACCGCTGACGCCCATCGAG	1407
Db	382	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal	401
Qy	1408	CTGCCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGTGGCAAGCTGAAC	1467
Db	402	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	421
Qy	1468	TGGGCGAGCGCATCTACCCGCGCATCAAGTGGCGGAGCTGTGCAAGCTGCTGGCGGC	1527
Db	422	TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly	441
Qy	1528	GCCAAAGCCCTGACCGACATCGTGGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAG	1587
Db	442	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu	461
Qy	1588	AACCGGAGATCTCGCGGAGCCCGTGCACCGCTGTACTACGACCCCGCAGCAAGCCTG	1647
Db	462	AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu	481
Qy	1648	GTGGCGAGATCCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1707
Db	482	IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro	501
Qy	1708	TTCAGAAGACCTGAAGACCGCAAGTACGCCAAGATCGCACCGCGCCACACACACGAGCTG	1767
Db	502	PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal	521
Qy	1768	AAGCAGCTGACCGAGGCGCTGCGAGAAGATCGCCATCGAGAGCATCGTGATCTGGGCAAG	1827
Db	522	LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTyrGlyLys	541
Qy	1828	ACCCCGAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGAGCTGGTGGACGACTAC	1887
Db	542	ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr	561
Qy	1888	TGGCAGCGCACCTGGAGTCCCGGAGTGGAGTTCGTGAACACCCCGCGCTGTGTGAAGCTG	1947
Db	562	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu	581
Qy	1948	TGTTACCGAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACTGTGGAGCGGCC	2007
Db	582	TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla	601
Qy	2008	GCCAAACCGGAGACCAAGATCGCGAGCGCGCTGCTGACCGCGCGCGCGCGCGGAG	2067
Db	602	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys	621



Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
QY 1171 ATCTGGAGCCCTTCGCGCCCGCAACCCGAGATCGTGATCTACAGTACATGAGCAG 1230  
Db 321 ILeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
QY 1231 CTGTACCTGGGCGAGCACTGGAGATCGGCAGCACCGCCCGCAAGATCGAGGAGCTGCGC 1290  
Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuArg 360  
QY 1291 AAGCACTGCTGCGCTGGGCTTCAACACCCCGCAAGAGCAGCAGAGAGCCCGCC 1350  
Db 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
QY 1351 TTCTGTGGATGGCTACGAGCTGCACCCGAGAGTGGACCTGAGCCCATCGAGCTG 1410  
Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
QY 1411 CCCGAGAAGGAGAGCTGGACCGGTGAACGACATCCAGAACTGGTGGCAAGCTGAACCTGG 1470  
Db 401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
QY 1471 GCCAGCAGATCTACCCCGCATCAAGTGGCGGAGCTGCAAGCTGCTGCGCGCGCC 1530  
Db 421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLysLeuLeuArgGlyAla 440  
QY 1531 AAGCCCTGACGACATCGTCCCTGACCGAGAGGCGGAGCTGGAGCTGGCGGAGAAC 1590  
Db 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460  
QY 1591 CGCGAGATCTCGCGAGCGCGTGCACGGCTGTACTACGACCCCGCAAGAGCACTGGTG 1650  
Db 461 ArgGluIleLysGluProValHisGlyValTyrTrpAspProSerLysAspLeuIle 480  
QY 1651 GCCGAGATCCAGAGAGGCGCAGACAGTGGACCTTACAGATCTACAGAGCCCTTC 1710  
Db 481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpTrpTyrGlnIleTyrGlnGluGlnTyr 500  
QY 1711 AAGAACTGAAGACCGGAGTAGTACCAAGATCGCACCGCCCGCAACACGAGCTGAAG 1770  
Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
QY 1771 CAGCTGACCGAGGCGGTGCAGAAAGATCCCATGAGAGCATCGTGATCTGGGGCAAGACC 1830  
Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
QY 1831 CCAGAGTCCGCTGCGCCATCCAGAGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890  
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560  
QY 1891 CAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGG 1950  
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
QY 1951 TACGAGCTGGAAGAGGAGCCCATCATCGCGCCCGAGACCTTCTACGTGGAGCGCGCGCC 2010  
Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 2011 AACCGCAGACCAAGATCGGAGCGCGGTACTGTGACCGCGGCGCGCGAGAGATC 2070  
Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
QY 2071 GTGAGCTGACCGAGACCCACAGACGAGCTGCAGGCGATCCAGCTGGCCCGCTG 2130  
Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
QY 2131 CAGGACAGCGGAGGAGTGAACATCGTACCGACAGCAGTACGCGCTGGGATCATC 2190  
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
QY 2191 CAGCCCGAGCCGACAAAGAGGAGGAGCTGGTGAACAGATCATCGAGCAGCTGATC 2250  
Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

QY 2251 AAGAAGGAGAAGGTGTACCTGAGCTGGTGCCTCCGCCCAAGGGCATCGCGGCAACGAG 2310  
Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700  
QY 2311 CAGATCGACAAGCTGTGTAGCAAGGCGCATCCGCAAGGTGCTTCTCTGAGCGGCATCGAT 2370  
Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 3  
US-08-463-210-9  
; Sequence 9, Application US/08463210  
; Patent No. 6001977  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, Nancy T.  
; APPLICANT: GALLO, Robert C.  
; APPLICANT: WONG-STALL, Flosie  
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morgan & Finnegan, L.L.P.  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,210  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 06/693,866  
; FILING DATE: 23-JAN-1985  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 06/659,339  
; FILING DATE: 10-OCT-1984  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Serunian, Leslie A.  
; REGISTRATION NUMBER: 35,353  
; REFERENCE/DOCKET NUMBER: 2026-4193US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1015 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HTLV-III  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..1015  
; OTHER INFORMATION: /note= "pol protein of HTLV-III"  
US-08-463-210-9

Alignment Scores: 8.35e-212 Length: 1015  
Pred. No.: 3513.00 Matches: 658  
Score: 3513.00  
Percent Similarity: 94.27% Conservative: 33  
Best Local Similarity: 89.77% Mismatches: 26  
Query Match: 76.65% Indels: 16  
DB: 3 Gaps: 3

US-09-610-313B-30 (1-2469) x US-08-463-210-9 (1-1015)

QY 220 TTCTTCCGCGAGGACCTGGCCCTTCCCTCCAGGGCAAGGCCCGCGAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyIysAlaArgGluPheSerSerGluGln 20  
QY 268 -----CCACGCGAGCAGAACCGCGCCCAACAGCCCGCCACAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
QY 304 CGCAGAGCTGCAGGTG-----CGCGGCGACAAACCCCGCAGCAGGGCGCGCGAGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
QY 358 CAGGGCAACCTTG-----AATCTCCCGCAGATCACCTGTGGCAGCGCCCTCGGTGAGC 411  
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
QY 412 ATCAAGGTGGCGCGCAGATCAAGAGGSCCTGTGGACACCGCGCGCGCAGCACCGTG 471  
Db 81 IleIysIleGlyGlnLeuIysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
QY 472 CTGAGGAGATGAGCTGCCCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGC 531  
Db 101 LeuGluMetSerLeuProGlyArgTrpIysProIysMetIleGlyGlyIleGlyGly 120  
QY 532 TTCTAAGGTGGCGCAGTACGACAGATCTGTATCGAGATCTCGCGCAGAAGGCCATC 591  
Db 121 PheIleIysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
QY 592 GGCAACGTGTGTATCGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651  
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
QY 652 CTGGGCTGCACCTCGAATCTCCCATCAGCCCATCAGACCGTGGCCCGTGAAGCTGAAG 711  
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CCGGCGATGACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCC 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCCCATCTCGGAGGAGATCGAGNAGGAGGCAAGATCACCAGATCGGCCCGCAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGlnGlyLysIleSerIysIleGlyProGlu 220  
QY 832 AACCCCTACAAACCCCGCTGTCGCATCAAGAAGAAGACAGACACCAAGTGGCGCAAG 891  
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTTGGAATCTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGC 951  
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCCACCCTCGCGCTGMAAGAGAGAGAGCTGACCGTGTGGAGCTGGCGGCAC 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTTCAGCTGCGCTGACGAGACTTCCGCAAGTACACCGCTTCCACCATCCCC 1071  
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACACGAGACCCCGCGATCCGCTACAGTACAAAGTGTGCCCCAGGCTGG 1131  
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGGACGCCAGCATCTCCAGAGCAGCATGACCAAGATCTGGAGCCCTTCCGCGCC 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLys 340  
QY 1192 CGCAACCCCGAGATCGGTATCTACAGTACATGACGACCTGTACGTGGGCGAGGACCTG 1251  
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

QY 1252 GAGATCGCGCGACCGCGCCAGATCGAGAGCTGCGCAAGCAGCCTGCTGGCTGGGCG 1311  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380  
QY 1312 TTCAACACCCCGCAACAAGACACCAAGAGAGCCCTTCTGTGTGATGGCTACGAG 1371  
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400  
QY 1372 CTGACACCCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACC 1431  
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1432 GTGAACACATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCGCG 1491  
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
QY 1492 ATCAAGGTGGCGCAGCTGTGCAAGCTGTGGCGCGCGCAAGGCGCTGACGACATCGTG 1551  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1552 CCCCTGACCGAGGAGCGCGAGCTGAGCTGGCCGAGAACCGCGAGATCTCTCGCGGAGCC 1611  
Db 461 ProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1612 GTGCAACCGCTGTACTAGCACCCGACGAGACCTGTGGCGCAGATCCAGACGAGGCC 1671  
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1672 CACGACAGTGGACCTACACAGATCTACGAGAGCCCTTCAAGAACCTGAACACCGGCAAG 1731  
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1732 TAGCCCAAGATGCGCACCGCCCAACACAGACGTGAAGCAGCTGACGAGCGCCGTGCAG 1791  
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1792 AAGATCCCATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATC 1851  
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProfile 560  
QY 1852 CAGAAGGAGACCTGGGAGACCTGTGTGACCGACTACTGCGAGGCGCACCTGGATCCCGAG 1911  
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1912 TGGAGGTTCGTGAACACCCCGCTGTAAGCTGTGTACCACTGGAGAGGAGAGGCC 1971  
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuLysGluPro 600  
QY 1972 ATCATCGCGCGCAGAGACTTCTAGTGGACGCGCGCCCGCAACCGCGAGACCAAGTCGC 2031  
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2032 AAGSCGCGCTACGTGACCGACCGCGCGCGCAGAGATCGTGAGCCTGACCGAGCACACC 2091  
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2092 AACCAGAGACCGAGCTGACGAGCCATCCAGCTGCGCCCTGCGAGCAGCGCGAGAGGTG 2151  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2152 AACATCGTGACGACGAGCATCGCCCTGGGCGATCATCGAGCCCGCGCAGACAGGCC 2211  
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2212 GAGAGCGAGCTGTGAACACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACTGT 2271  
Db 681 GluSerGlnLeuValAsnGlnIleIleGlnLeuIleLysLysValLysValTyrLeu 700  
QY 2272 AGCTGGGTGCGCGCCCAAGAGGCGCATCGCGCGCAACAGAGCAGATCGCAAGCTGCTGAGC 2331  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2332 AAGGGCATCCGCAAGGTGTGTTCTTCTGAGCGGCATCGAT 2370

```
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
|||||
RESULT 4
US-09-124-900-3
; Sequence 3, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andrea
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: PREDL, Renate
; APPLICANT: SCHMATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112P
; CURRENT APPLICATION NUMBER: US/09/124, 900
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: PCT/EP95/01481
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-124-900-3

Alignment Scores:
Pred. No.: 8,35e-212 Length: 1015
Score: 3513.00 Matches: 658
Percent Similarity: 94.27% Conservative: 33
Best Local Similarity: 89.77% Mismatches: 26
Query Match: 76.65% Indels: 16
DB: 3 Gaps: 3

US-09-610-313b-30 (1-2469) x US-09-124-900-3 (1-1015)
Qy 220 TTCTCCGCGAGGACCTGGCTTCCGCCAGGAGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyIleArgGluPheSerSerGluGln 20
Qy 268 -----CCGAGCGAGCAGAACCGCGCCCAACAGCCCCACCGAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
Qy 304 CGGAGCTGCGAGTG-----CGGCGGACACCCCGCGAGCGAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyIleArgAspAsnSerProSerGluAlaGlyAlaAspArg 60
Qy 358 CAGGCGACCCCTG-----AACTTCCCGACATCACCTGTGGCAGCGCCCTGTGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnProGlnIleThrLeuTrpGlnArgProLeuValThr 80
Qy 412 ATCAAGTGGCGCGCCAGATCAAGAGGCGCTGTGTGACACCGCGCGCGCGACCGCTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
Qy 472 CTGAGAGATGAGCTGCGCGGAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyIleArgTrpLysProLysMetIleGlyGlyIleGly 120
Qy 532 TTCAATCAAGTGGCGCGAGTACGACCATCTGATCGAGATCTCGCGCAAGAGCGGCATC 591
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluLeuIleCysGlyHisLysAlaIle 140
Qy 592 GGCACCGTGTGATCGCGCGCCACCCCGCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651
|||||
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
Qy 652 CTGGGCTGCACCTGAACCTTCCCATCAGCCCATCAGAGACCGTGCCTGGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
Qy 712 CCGGCGATGAGCGCCCAAGGTGAAGCAGTGGCCCTCAGCAGGAGAGATCAAGAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
Qy 772 CTGACCGCATCTCGAGGAGATGGAGAAGAGGAGGCAAGATCACCAAGATCGGCCCGC 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluLysIleSerLysIleGlyProGlu 220
Qy 832 AACCCCTACAACACCCCGCTTTCGCATCAAGAAGAAGACAGCAGCAGTGGCGCAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240
Qy 892 CTGGTGGACTTCGCGGAGCTGAACAAGCGCACCCAGGACTCTTGGGAGGTGCAGCTGGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
Qy 952 ATCCCGCACCCCGCGCTGAAGAAGAAGAGAGCGTGACCGTGTGAGACGTGGCGGAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280
Qy 1012 GCCTACTTCAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTCCACATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
Qy 1072 AGCATCAACAACAGAGACCCCGCATCCGCTACAGTACAAACGTGTGCCCGAGGCTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
Qy 1132 AGGGGAGCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGCCCTTCCGGGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
Qy 1192 CGCAACCCCGAGATCGTGATCTACACAGTACATGACGACCTGTACGTGGCGACGACCTG 1251
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeu 360
Qy 1252 GAGATCGCGCAGCACCGCGCCCAAGATCGAGGAGCTCGCAAGCACCTGTGCTGGCGGC 1311
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380
Qy 1312 TTCACACCCCGCAGCAAGAGCAGCAGAGGAGCCCGCTTCTGTGGATGGCTACGAG 1371
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400
Qy 1372 CTCACCCCGCAGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGCAGAGGAGCTGGACC 1431
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
Qy 1432 GTCAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCCAGCGCAGATCTACCCCGC 1491
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
Qy 1492 ATCAAGTGGCGCGCATGTGCAAGCTGTGCGCGCGCGCCAGCGCCCTGACCGCATCGTG 1551
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
Qy 1552 CCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGCGCGCGCCAGCGCCCTGACCGAGCC 1611
Db 461 ProLeuThrGluGluAlaGluLeuLeuAlaGluAsnArgGluLeuLysGluPro 480
Qy 1612 GTGCACCGGTGTACTACGACCCCAAGAGCCTGTGTGGCGCGAGATCTCAAGAGCGAGGC 1671
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
Qy 1672 CAGCAGCAGTGGACCTTACAGATCTACAGGAGCCCTTCAAGACCTGAGACCGCGCAAG 1731
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
```

```
QY 1732 TACGCCAAGATCCGACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGCGCGCTGCAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1792 AAGATCGCCATCGAGAGCATCGTATCTGGGGCAAGACCCCAAGTCCGCTGCCCATC 1851
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1852 CAGAAGAGACTGGGAGACCTGTGGACCGACTACTGCGACCGACCTGATCCCGAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTIpIleProGlu 580
QY 1912 TGGGAGTTCTGAACACCCCGCCCTGTGAAGCTGTGTACCTGAGTACCGAGAGAGGCC 1971
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1972 ATCATCGCGCGGACGACCTTCTACGTGGACGCGCGCCCAACCGGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2032 AAGCGCGCTACGTGACCGACCGCGCGCGGAGAGATCGTGAGCTTGACCGACACACC 2091
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2092 AACCAAGAGCGAGCTGAGCCATCCAGCTGCGCCCTGCGAGGACGCGGAGCGAGGTG 2151
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2152 AACATCGTGACGAGCAGCAGTACGCTCGGCGCATCCAGGCGCCCGACCAAGAGC 2211
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2212 GAGAGCGAGCTGTGAACACGATCATCCAGCAGCTGATCAAGAGGAGAGTGTACCTG 2271
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2272 AGCTGGGTGCGCGCCCAAGGCGCATCGGCGGCAACGAGCAGATCGACAAGCTGTGAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2332 AAGGCGATCCCGAAGGTCTGTCTCTGCGAGCGCATCGAT 2370
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

```
RESULT 5
US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAI, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESS: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,028
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
```

```
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1015
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
US-08-463-028-9
```

```
Alignment Scores:
Pred. No.: 8,35e-212 Length: 1015
Score: 3513.00 Matches: 658
Percent Similarity: 94.27% Conservative: 33
Best Local Similarity: 89.77% Mismatches: 26
Query Match: 76.65% Indels: 16
DB: 4 Gaps: 3

US-09-610-313B-30 (1-2469) x US-08-463-028-9 (1-1015)
QY 220 TTCCTCCGAGGACCTGGCTTCCCTCCCGAGGCAAGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 268 -----CCGAGCGAGGAGAACCGCGCCCAACAGCCCGCCACCGCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGCGGACAAACCCCGCAGCGCGCGCGCGCGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGCGACCCCTG-----AACTTCCCGACATCACCTGTGGCAGCGCCCTGTGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGTGGCGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGCGCGCGCGC 471
Db 81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100
QY 472 CTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
QY 532 TTCATCAAGTGGCGCGCAGTACGACAGATCTGTGATCGAGATCTGCGGCAAGAGCCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGCGCGCCCGCCCGTGAACATCATCGCGCGCGCAACATCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCTGAACTTCCCATCATGAGCCCGCGCGCGCGCGCGCGCGCGCGCGC 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
```



QY 712 CCCGCCATGACGCCCCCAAGGTTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCC 771  
DB 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACGCCCATCTCGAGGAGATGGAAGAGGGCAAGATCACCAGATCGGCCCGAG 831  
DB 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTACAACACCCCGCTTCCCATCAAGNAGAGACACCAAGTGGCGCAAG 891  
DB 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCCCGAGCTGAACAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGC 951  
DB 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCACCCCGCCGCTGAAGAAGAAGAGCGTGCCTGTCGAGCGTGGCGAC 1011  
DB 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTCAGCTGCCCTGGAGGAGACTTCCGCAAGTACACCGCTTCCACATCCCC 1071  
DB 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACAAGAGACCCCGCATCCGCTACCAAGTACCAACGTCTGCCCCAGGGCTGG 1131  
DB 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGGCACCCGACATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGGGCC 1191  
DB 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCGAGATCGTATCTACCAGTACATGAGACCTGTACCTGGGCGAGCGACTG 1251  
DB 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
QY 1252 GAGATCGGCAGACCCGCGCAAGATCGAGAGTGGCGAAGACCTGCTGCGCTGGGGC 1311  
DB 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuAspTrpGly 380  
QY 1312 TTCAACCCCGCAAGAACAGCACCAAGAGAGCCCTTCTCTGTGTGATGGGTACGAG 1371  
DB 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400  
QY 1372 CTGCACCCGACAGTGGACGCTGCAGCCCATCGAGTGGCGCCAGAGAGAGAGTGGACC 1431  
DB 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1432 GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAATGGGCGCCAGCATCTACCCCGGC 1491  
DB 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
QY 1492 ATCAAGTGGCCGACGTGTGCAAGCTGCTGCGGGCGCCAAAGCCCTGACCGACATCGTG 1551  
DB 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1552 CCCTGTACCGAGGCGGAGCTGGAGCTGGCCGAGAACCGGAGATCCTGCGCGAGCCC 1611  
DB 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1612 GTGCACCGGCTGTACTACGACCCCGCAAGACCTGGTGGCGCCAGATCCAGAGAGAGGC 1671  
DB 481 ValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1672 CACGACAGTGGACCTACAGATCTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731  
DB 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1732 TAGCCCAAGATGCCACCGCCCAACCAACGACGTGAAGCAGTGCAGCGGCGGTGCAG 1791  
DB 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1792 AAGATCGCCATGGAGATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATC 1851

DB 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1852 CAGAAGAGACCTGGAGAGACCTGGTGGACCGACTACTGGCAGGCGCCACCTGGATCCCCGAG 1911  
DB 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1912 TGGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGGTACCAGCTGGAGAAGAGGCC 1971  
DB 581 TrpGluPheValAsnThrProProLeuValLysLeuTyrTrpGlnLeuGluLysGluPro 600  
QY 1972 ATCATCGGCGCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGC 2031  
DB 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2032 AAGCGCGGTACGTGACCGGCGCGCGCGCGCAGCAAGATCGTGGCTGACCGAGACCCACC 2091  
DB 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2092 AACCAAGAAGACCGAGTGCAGGCGCATCCAGCTGGCTGGCATCATCCAGCCCGCAGAGCGGTG 2151  
DB 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2152 AACATCGTGACCGACAGCCAGTACGCTGGGCATCATCCAGCCCGCAGCCCGCAAGAGC 2211  
DB 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2212 GACAGCGAGCTGTGTGAACCGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTG 2271  
DB 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700  
QY 2272 AGCTGGTGGCCCGCCCAAGGCGCATCGGCGCAAGCAGCAGATCGCAAGCTGGTGAGC 2331  
DB 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2332 AAGGCGATCCGCAAGTGTCTTCTGGAGCGGCATCGAT 2370  
DB 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 6  
US-07-743-357-2  
; Sequence 2, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900

TELFAX: (613) 237-0045  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1016 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 FRAGMENT TYPE: internal  
 ORIGINAL SOURCE:  
 ORGANISM: Human immunodeficiency virus type 1  
 STRAIN: BH102  
 US-07-743-357-2

## Alignment Scores:

Pred. No.: 8,36e-212 Length: 1016  
 Score: 3513.00 Matches: 658  
 Percent Similarity: 94.27% Conservative: 33  
 Best Local Similarity: 89.77% Mismatches: 26  
 Query Match: 76.65% Indels: 16  
 DB: 2 Gaps: 3

US-09-610-313B-30 (1-2469) x US-07-743-357-2 (1-1016)

220 TTCTTCGCGAGGACCTGGCTTCCCGAGGCGAAGGCCCGCGAGTTC----- 267  
 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
 268 -----CCGAGCGAGCAGAACCGCGGCAACAGCCCGCGAGGCGCGCGC 303  
 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
 304 CGCGAGCTGCAGGTG-----CGCGGCGACAACCCCGCAGCGCGCGCGCGCGCGC 357  
 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
 358 CAGGCGACCCCTG-----AATCTCCCGCAGATCACCTGTGGCAGCGCCCTCGTGAGC 411  
 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
 412 ATCAAGTGGCGCGCGAGATCAAGAGCGCTGTGGACACCGCGCGCGCGCGAGCACCGTG 471  
 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
 472 CTGAGGAGATGACCTGCGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531  
 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyGly 120  
 532 TTCAATCAAGTGGCGCGAGTACGACCATCTGTGATCGAGATCTCGCGCAAGAAGCCATC 591  
 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
 592 GGACACCTGTGATCGC 651  
 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuGln 160  
 652 CTGGGCTGCACTTCCCACTTCCCATCAGCCCATCAGACCGCGCGCGCGCGCGCGCGCGC 711  
 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
 712 CCGGCGATGAGC 771  
 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
 772 CTGACCCCGCATCTCGGAGGAGATGAGAGAGGCGCGAAGATCACCAAGATCGGCGCGCGC 831  
 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
 832 AACCCCTACACACCCCGGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

QY 892 CTGCTGGACTTCCGCGAGCTGAACAAGCGGACCCAGGACTTCTTGGAGGTGCGAGTGGGC 951  
 DB |||||  
 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
 QY 952 ATCCCCCAACCCCGCGCTGAAGAAGAAGAGCGTGACCGTCTGCGACTGGCGCGAC 1011  
 DB |||||  
 261 IleProHisProAlaGlyLeuLysLysSerValThrValLeuAspValGlyAsp 280  
 QY 1012 GCCTACTTCAAGCGTGGCGCGCTTCCGAGGAGACTTCCGCAAGTACACCGCTTCCACATCCCC 1071  
 DB |||||  
 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
 QY 1072 AGCATCAACAACGAGACCCCGCGCATCCGCTACAGTACACAGCTGCTGCCCGAGGCTGG 1131  
 DB |||||  
 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
 QY 1132 AAGGCGACCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCC 1191  
 DB |||||  
 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
 QY 1192 CGCAACCCCGAGATCGTGATCTACCGAGTACATGGACGACCTGTAGTGGGCGAGCGACCTG 1251  
 DB |||||  
 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
 QY 1252 GAGATCGCGCGAGCACCGCGCAAGATCGAGAGCTGCGCAAGACCTCTCGCTGGTGGGC 1311  
 DB |||||  
 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380  
 QY 1312 TTCAACACCCCGCAAGAGCACCAAGAGAGCCCGCTTCTGTGGATGGCTACCG 1371  
 DB |||||  
 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400  
 QY 1372 CTGACACCCCGCAAGTGGACCGCTGCGACCCATCGAGCTGCCGAGAGAGAGAGTGGAC 1431  
 DB |||||  
 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
 QY 1432 GTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCGAGCGAGATCTACCCCGC 1491  
 DB |||||  
 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
 QY 1492 ATCAAGTGGCGCGAGCTGCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1551  
 DB |||||  
 441 IleLysValArgGlnLeuLysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
 QY 1552 CCCCTGACCGAGGAGCGCGAGCTGGAGCTGCCGAGAACCGCGAGATCTCTCGCGAGCGC 1611  
 DB |||||  
 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
 QY 1612 GTGACGCGGTGTACTACGACCCCGAGACCTGTGGCGCGAGATCCAGAAGCAGGCG 1671  
 DB |||||  
 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
 QY 1672 CACGACGAGTGGACCTACAGATCTACGAGAGCGCTTCAAGACCTCAAGACCGCGCAAG 1731  
 DB |||||  
 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
 QY 1732 TACGCCAAGATGGC 1791  
 DB |||||  
 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
 QY 1792 AAGATCCCGCATCGAGACATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATC 1851  
 DB |||||  
 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
 QY 1852 CAGNAGGACCTTGGGAGACCTGTGACCGAGCTACTTGGCAGGCGCGCGCGCGCGCGCGC 1911  
 DB |||||  
 561 GlnLysGluThrTrpGluThrTrpTrpThrGlyTyrGlnAlaThrTrpIleProGlu 580  
 QY 1912 TGGGAGTTCTGAAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1971  
 DB |||||  
 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600  
 QY 1972 ATCATCGCGCGCGAGACCTTCTTACGTGAGCGCGCGCGCGCGCGCGCGCGCGCGC 2031



```
QY 1528 GCCAAGCCCTGACCGACATGTCCTCCCTGACCGAGAGCCGAGCTGAGCTGGCCGAG 1587
Db 441 ThrLysAlaLeuThrGluValProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1588 AACCGCAGATCCGCGCGCGCTGACGGCTGTACTACGACCCCGACGAGCCTG 1647
Db 461 AsnArgGluLeuLeuGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1648 GTGCCCGAGATCCAGAGCAGCGCCACGACAGTGGACCTACACAGATCTACGAGAGCCC 1707
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
QY 1708 TTCAGAACCTGAAGACCGGCAATGACGCCAAGATGCGCACCGCCACACCAACGAGCTG 1767
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGCGCGTGCAGAAGATGCCATGAGAGCATCGTGAATCTGGGCAAG 1827
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTyrGlyLys 540
QY 1828 ACCCCCAAGTTCGCCCTGCCATCCAGAAGAGACCTGGGAGACCTGGTGACCGACTAC 1887
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTyrGluAlaTyrTyrGluTyr 560
QY 1888 TGGCAGCCACCTGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTG 1947
Db 561 TrpGlnAlaThrTyrIleProGluTyrGluPheValAsnThrProLeuValLysLeu 580
QY 1948 TGGTACAGCTGGAGAGAGCCCATCATCGGCGCGGAGCTTCTACGTGGAGCGGCC 2007
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2008 GCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGCGCGCGCAGAAG 2067
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
QY 2068 ATCGTGAAGTCCGACGACCAACCAACGAGACCGAGTGCAGGCGCATCCAGCTGGCC 2127
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2128 CTGCAGACAGCGCGCAGGAGTGAACATCGTGACCGACGACGAGTACGCGCGCTGGGCATC 2187
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGGCCCCAGCCGACAGAGCGAGCGAGCTGGTGAACACGATCATCGACGAGCTG 2247
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCACCAAGGGCATCGCGGCAAC 2307
Db 681 IleLysLysGluLysValTyrLeuAlaTyrValProAlaHisLysGlyIleGlyAsn 700
QY 2308 GAGCAGATCGACAGCTGGTGAGCAAGGGCATCGCAAGGTGCTGCTTCCTGGAGCGGCATC 2367
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db 721 Asp 721
```

## RESULT 8

```
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718,096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
```

```
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-718-096-17
```

```
Alignment Scores:
Pred. No.: 1.28e-211 Length: 1003
Score: 3510.00 Matches: 656
Percent Similarity: 95.70% Conservative: 34
Best Local Similarity: 90.98% Mismatches: 27
Query Match: 76.59% Indels: 4
DB: 4 Gaps: 2
```

US-09-610-313B-30 (1-2469) x US-09-718-096-17 (1-1003)

```
QY 220 TTCTTCGCGAGGACCTGGGCTTCCCCAGGGGAGAGCCCGGAGTTCCCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCCCAACAGCCACCCAGCCGCGAGCTGCAGGTG-----CGCGGCGACAAACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCAGGCGCGCGCGCGAGCGCGAGCCAGCCCTG-----AACTTCCCCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGGCCCCCTGGTGTGAGCATCAAGTGGCGGCGCGAGATCAAGAGGCGCCCTGTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGAGCACCGCTGCGAGGAGATGAGCTGCCCGGCAAGTGAAGAGCCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCGCAGTACGACCATCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTGGCGCAAGAGGCGCATCGGCACCGCTGTGTGATCGGCCCGCCCGCTGAAATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGCTGACCTGACCTTCCCATCTCCCATCAGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGGCCGTGAAGCTGAAGCCCGGATGAGCGGCCCGCCCAAGGTGAAGCAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCGCTGACCGCCCATCTGCGAGGAGATGGAGAGAGGCGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCGCGCGAGAACCCCTACAAACACCCCGCTGCGCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTCAACCAAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGGAGGTGACGTGGGCATCCCCACACCCCGCGGCTCTGAAGAAGAAGAGAGC 987
```

Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysSer 260  
QY 988 GTGACCGTGTGAGCGTGGCGAGCCCTACTTACGCGTGCCCTGAGCAGGACTTCCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
QY 1048 AAGTACACCGCTTCCACATCCCGCAGCATCAACAGAGACCCCGCATCCGCTACCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGlnThrProGlyIleArgTyrGln 300  
QY 1108 TACAACGTGTGCCCCAGGCGCTGGAAGGCGAGCGCCAGCATCTTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320  
QY 1168 AAGATCTGTGAGCGCTTCCCGCCCGCAACCCCGAGATCGTGATCTACCATGTCATGGAC 1227  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
QY 1228 GACCTGTACGTGGCGAGCGACCTGGAGATCGGCAGCAGCCCGCCAGATCGAGGAGCTG 1287  
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1288 CGCAAGCACCTGTGCTCGGCGCTTCAACACCCCGCACAAAGACACCAAGAGAGGCC 1347  
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrTrpProAspLysLysHisGlnLysGluPro 380  
QY 1348 CCTTCTGTGGATGGCTACGAGTGCACCCCGACCAAGTGGACCGTGCAGCCCATCGAG 1407  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
QY 1408 CTCCCGAGAGAGAGCTGCAGCGTGAACAGCATCCAGAGCTGTGGGCAAGCTGAAC 1467  
Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1468 TGGCGCCAGCAGATCTACCCCGCGCATCAAGGTGGCGCGAGCTGTGCAAGCTGCTGCGCGC 1527  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
QY 1528 GCGAAGCCCTGACCGACATCGTGCCTGACCGAGAGCGCGAGCTGGAGCTGGCGCGAG 1587  
Db 441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
QY 1588 AACCGCAGATCTTCGCGAGCGCGTGCACGGGTGTACTACGACCCCGCAGCAGGAGCTG 1647  
Db 461 AsnArgGluIleLysGluProValHisGlyValTyrAspProSerLysAspLeu 480  
QY 1648 GTGCGCCAGATCCAGAGCAGGCGCACGACAGTGGACCTTACCAGATCTTACCAGAGCCC 1707  
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500  
QY 1708 TTCAAGAACTGGAAGCCGGAAGTAGCCCAAGATGCGCACCGCCCAACACCAACGAGCTG 1767  
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520  
QY 1768 AAGCAGCTGACCGAGCGCGTGCAGAGATCGCAATCGCATCGAGAGCATCGTGATCTGGGCAAG 1827  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
QY 1828 ACCCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTTGGGAGACCTGGTGACCGACTAC 1887  
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTyr 560  
QY 1888 TGGCAGCCCACTGGATCCCGAGTGGGAGTTGTTGAACACCCCGCCCTGGTGAAGCTG 1947  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
QY 1948 TGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGCC 2007  
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600  
QY 2008 GCCAACCGCAGAGCAAGATCGGAGCGCGGTACGTGACCGAGCGGCGCGCGAGAG 2067

Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620  
QY 2068 ATCGTGAGCTGACCGAGACCCACCAACAGACGAGCTGAGAGCTGAGAGCTGAGCTGGCC 2127  
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
QY 2128 CTGACGAGCAGCGCGCAGCGAGGTGAACATCGTGACCGCAGCAGCAGTACGCCCTGGGCATC 2187  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
QY 2188 ATCCAGGCCAGCCCGACCAAGAGCAGAGCAGCTGGTGAACCCAGATCATCGAGCAGCTG 2247  
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
QY 2248 ATCAAGAGAGAGAGGTGTACTGAGCTGGTGCCGCCACCAAGGCGCATCGCGGCAAC 2307  
Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
QY 2308 GAGCAGATCGACAAGCTGGTGAAGAGGCGCATCCGCAAGGTGCTGTCTTGGAGCGGCATC 2367  
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
QY 2368 GAT 2370  
Db 721 Asp 721  
RESULT 9  
US-07-743-357-5  
; Sequence 5, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1016 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: BRU

US-07-743-357-5

## Alignment Scores:

Pred. No.: 1.98e-211 Length: 1016  
 Score: 3507.00 Matches: 656  
 Percent Similarity: 94.41% Conservative: 36  
 Best Local Similarity: 89.50% Mismatches: 25  
 Query Match: 76.52% Indels: 16  
 DB: 2 Gaps: 3

US-09-610-313B-30 (1-2469) x US-07-743-357-5 (1-1016)

Qy 220 TTCTTCGGCAGGACCTGGCCCTTCCCGAGGGCAAGGCCCGCGAGTTC----- 267  
 Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
 268 -----CCCGAGCAGAGAACCGCGCAACAGACAGCCCGCAGCAGC 303  
 Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
 Qy 304 CGCGAGCTGCAGGTG-----CGCGGCGACAACCCCGCAGCGAGCGCGCGCGAGCGC 357  
 Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArg 60  
 Qy 358 CAGGGCACCCCTG-----AACTTCCCGCAGATCACCCTGTGGCAGCGCCCTCGTGAGC 411  
 Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
 Qy 412 ATCAAGGTGGCGCGCAGATCAAGGAGCGCTGTGGACACCGCGCGCGCAGCAGCACCGTG 471  
 Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
 Qy 472 CTGAGGAGATGAGCCCTGCGCGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGGC 531  
 Db 101 LeuGluGlnMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyGly 120  
 Qy 532 TTGATCAAGGTGCGCAGTAGCAGATCCTGATCGAGATCTCGCGCAAGAGCGCCATC 591  
 Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
 Qy 592 GGACCGGTGTGATCGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGCACCCAG 651  
 Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
 Qy 652 CTGGGCTGCACCTCGAATCTCCCATCAGCCCATCAGACCGTGGCCCGTGAAGCTGAAG 711  
 Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
 Qy 712 CCCGGCATGAGCGCCCGCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCC 771  
 Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
 Qy 772 CTGACCGCCATCTCGGAGGATGAGAGAGAGGCGCAAGATCACCAGATCGCGCCCGAG 831  
 Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
 Qy 832 AACCCCTTACACACCCCGTCTCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
 Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
 Qy 892 CTGTGTGACTTCCCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGC 951  
 Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
 Qy 952 ATCCCGCCACCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
 Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
 Qy 1012 GCCTACTTTCAGCTGCGCGAGGACTTCCGCAAGTATCACCCTTCACCATCCCC 1071  
 Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
 Qy 1072 AGCATCAACACGAGACCCCGCGCATCCGCTACAGTACACAGTGTGCCCCAGGGCTGG 1131

Db 301 SerIleAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
 Qy 1132 AAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGGCC 1191  
 Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340  
 Qy 1192 CGCAACCCCGAGATCGTGATCTACAGTACATGACGACCTGTAGCTGGGAGGAGACCTG 1251  
 Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
 Qy 1252 GAGATCGCGCAGCAGCGCCCAAGATCGAGAGCTGCGCAAGCAGCTGCTGGCTGGGGC 1311  
 Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380  
 Qy 1312 TTCAACACCCCGCAGCAAGAGCAGCAGAGAGCGCCCTCTCTGTGTGATGGGTACTACGAG 1371  
 Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400  
 Qy 1372 CTGCACCCCGCAAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGTGGACC 1431  
 Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
 Qy 1432 GTGAACACATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGCAGCCAGATCTACCCCGC 1491  
 Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
 Qy 1492 ATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACGACATCGTG 1551  
 Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
 Qy 1552 CCCGTGACCGAGGAGCGAGCTGTGGCCGAGAACCGCGAGATCTCTCGCGAGGCC 1611  
 Db 461 ProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
 Qy 1612 GTGCACCGGTGTACTAGACCCCGCAGCAGACCTGTGGCGCAGATGCCAAGCAGCGGC 1671  
 Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
 Qy 1672 CACGACAGTGGAGCTACACAGATCTACGAGAGCGCTTCAAGAACCTGAAGACCGCGAAG 1731  
 Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnLysProPheLysAsnLeuLysThrGlyLys 520  
 Qy 1732 TAGCCCAAGTGGCCACCGCCCAACCAAGCTGAAGCAGCTGACGAGCGCGCTGCGAG 1791  
 Db 521 TyrAlaArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
 Qy 1792 AAGATCCCATGAGCAGCATCTGTGGGGCAAGACCCCAAGTTCGGCTCCGCATC 1851  
 Db 541 LysIleThrThrGluSerIleValIleTrpLysThrProLysPheLysLeuProIle 560  
 Qy 1852 CAGAAGGAGACCTGGGAGACCTGTGGACGAGTACTTGGCAGGCGCACCTGGATCCCGAG 1911  
 Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
 Qy 1912 TGGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGTACCAAGCTGGAGAGAGAGGCC 1971  
 Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuLysGluPro 600  
 Qy 1972 ATCATCGCGCGGAGACCTTCTAGCTGACCGCGCGCGCCCAACCGCGAGACCAAGATCGC 2031  
 Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620  
 Qy 2032 AAGGCGCGCTAGCTGAGCAGCGCGCGCGCGCAGAGATCGTGAGCTGACCGAGACCAACC 2091  
 Db 621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValThrLeuThrAspThrThr 640  
 Qy 2092 AACCAAGAGCCGAGCTGCAGGCCATCCAGCTGGCCCTGCGAGGACAGCGGAGCGAGGTG 2151  
 Db 641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
 Qy 2152 AACATCTGACCGCAGCCAGTACGCCCTTGGGCATCATCCAGGCCCGCGCAGCAAGGC 2211



Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleGlnAlaGlnProAspLysSer 680  
 QY 2212 GAGAGCGAGTGTGTAACCATCATCGACAGCTGATCAAGAGGAGAGGTGTACCTG 2271  
 Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGluLysValTyrLeu 700  
 QY 2272 AGCTGGGTGCGCGCCCAACAGGGCATCGCGGCAACGAGCAGATCGCAAGCTGTGGTGAGC 2331  
 Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSer 720  
 QY 2332 AAGCGCATCCGCAAGTGTGTTCTCGACGGCATCGAT 2370  
 Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

## RESULT 10

US-07-743-357-3  
 ; Sequence 3, Application US/07743357  
 ; Patent No. 5858646  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kang, Yong C.  
 ; TITLE OF INVENTION: Polypeptide having immunological  
 ; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: KIRBY EADES GALE BAKER  
 ; STREET: Box 3432, Station D  
 ; CITY: Ottawa  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: K1M 1H8  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/743,357  
 ; FILING DATE: 21-AUG-1991  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/CA90/00062  
 ; FILING DATE: 23-FEB-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gale, Edwin J.  
 ; REGISTRATION NUMBER: 28,584  
 ; REFERENCE/DOCKET NUMBER: 30924-2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (613) 237-6900  
 ; TELEFAX: (613) 237-0045  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1016 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; FRAGMENT TYPE: internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Human immunodeficiency virus type 1  
 ; STRAIN: BH5  
 ; US-07-743-357-3

Alignment Scores:  
 Pred. No.: 2,29e-211 Length: 1016  
 Score: 3506.00 Matches: 656  
 Percent Similarity: 94.27% Conservative: 35  
 Best Local Similarity: 89.50% Mismatches: 26  
 Query Match: 76.50% Indels: 16  
 DB: 2 Gaps: 3

US-09-610-313B-30 (1-2469) x US-07-743-357-3 (1-1016)

QY 220 TTCTTCCGCGAGACCTGGCTTCCGCCAGGGCAAGCCCGCGAGTTC----- 267  
 Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
 QY 268 -----CCAGCGAGCAGAACCGCGCCCAACAGCCCCCACCAGC 303  
 Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
 QY 304 CGCGAGCTCCAGGTG-----CGCGCGGACAACCCCGCAGCGAGCGCGCGCGAGCGC 357  
 Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
 QY 358 CAGGCGACCCCTG-----AACTTCCCGCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGC 411  
 Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
 QY 412 ATCAAGGTGGCGCGCGCAGATCAAGGAGGCGCTCTCTGGACACCGCGCGCGAGCACCGGTG 471  
 Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100  
 QY 472 CTGGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCATCGCGCGCATCGCGCGC 531  
 Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120  
 QY 532 TTCTATCAAGGTGCGCGCAGTACGACCGATCTCTGATCGAGATCTCGCGCAAGAGGCCATC 591  
 Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
 QY 592 GGCACCGTGTGATCGCGCGCCACCCCGGTGAACATCATCGCGCGCGCAACATGCTGACCCAG 651  
 Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
 QY 652 CTGGGCTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAG 711  
 Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleIleuThrValProValLysLeuLys 180  
 QY 712 CCGGCATGGACCGCGCCCAAGGTGAAGCAGTGGCCCTGTACCGGAGGAGAGATCAAGGCC 771  
 Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
 QY 772 CTGACCGCATCTCGGAGGAGATGGAGAGGAGGGAAGATCATCCAGATCGCGCGCGCAG 831  
 Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
 QY 832 AACCCCTACAACACCCCGTGTTCGCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG 891  
 Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
 QY 892 CTGGTGGACTTCGCGAGCTGAAACAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGC 951  
 Db 241 LeuValAspPheArgGluLeuAsnArgArgThrGlnAspPheTrpGluValGlnLeuGly 260  
 QY 952 ATCCCGCACCCCGCGCTGAAGAAGAAGAGAGCGTGAACCTGCTGGAGCTGGGGCGAC 1011  
 Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
 QY 1012 GCCTACTTCAGCGTGCCTCGGACGAGACTTCGCAAGTACACCGCTTCCACCATCCCC 1071  
 Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
 QY 1072 AGCATCAACAACGAGACCCCGCGCATCCCGTACCAGTACCAACGTGTGCTGCCCGAGGTGG 1131  
 Db 301 SerIleAsnAsnGluThrProGlySerGlyTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
 QY 1132 AAGGGAGCCCGCAGCATCTTCCAGAGCAGCATCAACAGATCTGGAGCCCTTCCGCGCC 1191  
 Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340  
 QY 1192 CGCAACCCCGAGATCGTGATCTACCATGACGACCTGTACGTGGCGCAGCAGCTG 1251  
 Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
 QY 1252 GAGATCGGCCAGCACCCGCGCAAGAGTGGAGAGCTGGCAAGCACCTGTGCTGGCGGC 1311



```
QY 358 CAGGGCACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGCGCCCTCGGTGAGC 411
Db      |||
QY 61  GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
Db      |||
QY 412 ATCAAGTGGCGGCAGATCAAGAGGCCCTGCTGGACACCGCGCCGACGACACCGTG 471
Db      |||
QY 81  IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGC 531
Db      |||
QY 101 LeuGluGlnMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGly 120
QY 532 TTCAATCAAGTGGCGGCAGTACGACAGATCTCTGATCAGATCTCGCGCAAGAGGCCATC 591
Db      |||
QY 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGACCCGTGTGATCGGCCCCACCCCGTGAACATCATCGCGCGCGCAACATGCTGACCCAG 651
Db      |||
QY 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCTGAACTTCCCATCAGCCCATCGAGACCTGCGCGTGAAGCTGAAG 711
Db      |||
QY 161 IleGlyCysThrLeuAsnPheProIleSerProIleGlnThrValProValLysLeuLys 180
QY 712 CCGGCATGACCGCCGCCAAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCC 771
Db      |||
QY 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCCCATCTCGGAGGAGATGGAGAAGAGGGCAAGATCATCCAAAGATCGGCCCGGAG 831
Db      |||
QY 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
QY 832 AACCCCTACAAACCCCGCTTCGCCATCAAGAAGAAGCAGACCAAGTGCSCGCAAG 891
Db      |||
QY 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGGTGAATCTCCGCGAGCTGAACAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGC 951
Db      |||
QY 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCCCACCCCGCCCTGAAGAAGAAGAAGCGTGACCGTGTGGACGTGGCGGCAC 1011
Db      |||
QY 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTTCACTGCGCCCTGGAGGAGACTTCGCGAAGTACACCGCTTCCACCATCCCC 1071
Db      |||
QY 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACACGAGACCCCGGCATCCGCTACCAAGTACAACTGTCTGCCCGCCAGGGCTGG 1131
Db      |||
QY 301 SerIleAsnAsnGlnThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGGCACCCCGACATCTTCCAGAGCAGATGATCAACAGATCTCGAGAGCCCTTCCGCGCC 1191
Db      |||
QY 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340
QY 1192 CGCAACCCCGAGATCGTGTACTACCAATGACACCTGTACGTGGCGCAGCGACCTG 1251
Db      |||
QY 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
QY 1252 GAGATCGGCAGCACCCGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGCGCGTGGGGC 1311
Db      |||
QY 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
QY 1312 TTCACACCCCGCAAGACGACCAAGAGGCCCTTCTGTTGGATGGGCTAGCAG 1371
Db      |||
QY 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1372 CTGCACCCCGACAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGTGGACC 1431
Db      |||
QY 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1432 GTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACCTGGGCGCACGACATCTACCCCGGC 1491
```

```
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
QY 1492 ATCAAGTGGCGGCAGCTGTGCAAGCTGCGCGCGCCCAAGCCCTGACCGACATCGTG 1551
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1552 CCCTCACCAGGAGGCGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGCGAGCCC 1611
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1612 GTGCACGGGTGTACTACGACCCCAAGACCTGCTGGCCGAGATCCAGAGACGAGGC 1671
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1672 CACGACCACTGGACCTTACCAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnLysProPheLysAsnLeuLysThrGlyLys 520
QY 1732 TAGCCCAAGATGCGCACCCGCCACCAACGACGCTGAAGCAGCTGACCGAGGCGCTGCAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1792 AAGATCGCCATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCCCTGCCCATC 1851
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1852 CAGAAGGACCTGGGAGACCTGGTGACCCGACTACTGGCAGGCCACCTGGATCCCGGAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1912 TGGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTACAGCTGGAGAGAGGACCC 1971
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTyrTrpGlnLeuLysGluPro 600
QY 1972 ATCATCGGGCCGAGACCTTCTACGTGGAGCGCGCCGACCCGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGly 620
QY 2032 AAGCCCGCTTACGTGACCGACCGCGCGCGCAAGATCGTGAGCTGACCGAGACCCACC 2091
Db 621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValProLeuThrAsnThrThr 640
QY 2092 AACCAAGAAGACCGAGCTGCGAGCCATCCAGCTGGCCCTGCGAGACACGCGCAGCGGTG 2151
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2152 AACATCGTGACCGACGACGATACGCCCTGGGGCATCATCCAGGCCCGCCGACAGAGC 2211
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSer 680
QY 2212 GAGACGAGCTGGTGAACCAAGATCATCGAGCAGCTCATCAAGAGGAGAGGTGTACCTG 2271
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuLysGlnLysValTyrLeu 700
QY 2272 AGCTGGGTGCCCGCCACCAAGGCATCGCGCGCAACGAGCAGATCGCAAGCTGGTGAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2332 AAGGCATCGCAAGTGTCTTCTGGACGGCATCGAT 2370
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

## RESULT 12

US-07-743-357-10

; Sequence 10, Application US/07743357

; Patent No. 5858646

; GENERAL INFORMATION:

; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological

; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

ADDRESSEE: KIRBY EADES GALE BAKER  
 STREET: Box 3432, Station D  
 CITY: Ottawa  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: K1M 1H8  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/743,357  
 FILING DATE: 21-AUG-1991  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/CA90/00062  
 FILING DATE: 23-FEB-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gale, Edwin J.  
 REGISTRATION NUMBER: 28,584  
 REFERENCE/DOCKET NUMBER: 30924-2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (613) 237-6900  
 TELEFAX: (613) 237-0045  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1003 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 FRAGMENT TYPE: internal  
 ORIGINAL SOURCE:  
 ORGANISM: Human immunodeficiency virus type 1  
 STRAIN: ELI  
 US-07-743-357-10

Alignment Scores:  
 Pred. No.: 7,76e-211 Length: 1003  
 Score: 3497.50 Matches: 650  
 Percent Similarity: 95.83% Conservatives: 40  
 Best Local Similarity: 90.28% Mismatches: 27  
 Query Match: 76.31% Indels: 3  
 DB: 2 Gaps: 2

US-09-610-313B-30 (1-2469) x US-07-743-357-10 (1-1003)

Qy	571	ATCTGGCGCAAGAGGCGCATCGGCACCGTGTCTGATCGGCCCGCCCGTGAACATCATC	630
Db	121	IleCysGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle	140
Qy	631	GGCGCAACATGCTGACCCAGCTGGGTGACCTGAACCTTCCCATCAGCCCATCAG	690
Db	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
Qy	691	ACGTGCCCCGTGAAGTGAAGCCCGCATGAGCGGCCCAAGAGTGAAGAGTGCCTG	750
Db	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
Qy	751	ACGAGGAGAGATCAAGGCGCTGACCGCCATCTCGGAGGAGATGGAGAGAGGCGCAAG	810
Db	181	ThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys	200
Qy	811	ATCACCAAGATCGGCCCGGAGAACCTCTACAACACCCCGGTGTCGCGCATCAAGAAGAG	870
Db	201	IleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys	220
Qy	871	GACAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGGAC	930
Db	221	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	240
Qy	931	TTCTGGGAGGTGCGAGCTGGGCATCCCCACCCCGCGCTGAGAGAGAGAGAGCGGTG	990
Db	241	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	260
Qy	991	ACGTGCTGACGTGGCGGCGACCTACTTACGCGTGCCTCGGAGGAGGACTTCGCGAAG	1050
Db	261	ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys	280
Qy	1051	TACACCGCTTACCATCCCGAGCATCAACAAGAGAGACCCCGCGCATCCGCTACCAAGTAC	1110
Db	281	TyrThrAlaPheThrIleSerSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr	300
Qy	1111	ACGTGCTGCCCGCGGCTGAGGAGGAGGCGCCAGCATCTTCCAGAGCAGCATACCAAG	1170
Db	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
Qy	1171	ATCTGGAGCGCTTCGCGCGCGCAACCCCGAGATCGTATCTACCATGATACATGAGCAG	1230
Db	321	IleLeuGluProPheArgLysGlnAsnProGluMetValIleTyrGlnTyrMetAspAsp	340
Qy	1231	CTGTACCTGGCGCAGCACCTGGAGATCGGCAGCACCGCGCGCAAGATCGAGGAGCTGCGC	1290
Db	341	LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg	360
Qy	1291	AAGCACCTGCTGGCGCTTCCACCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1350
Db	361	GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro	380
Qy	1351	TTCTGTGATGGCTACGAGCTGCACCCCGAGAGTGGACCGCGAGAGAGAGAGAGAGAG	1410
Db	381	PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu	400
Qy	1411	CCGAG	1470
Db	401	ProGluLysGluSerTrpThrValAsnAspIleGlnAsnLeuValGluArgLeuAsnTrp	420
Qy	1471	GCAGAGCAGATCTACCCCGCGCATCAGGTGGCGCGAGCTGTGCAAGCTGTGCGCGGCGCC	1530
Db	421	AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuLysLysLeuLeuArgGlyThr	440
Qy	1531	AAGGCGCTGACGACATCGTGCCTGACCGAGGAGGCGCGAGCTGGAGCTGGCGGAGAG	1590
Db	441	LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGluAsn	460
Qy	1591	CGGAGATCTTGGCGGAGCGCGTGCACGGGTGTACTACGACCCCGAGAGAGAGAGAGAG	1650
Db	461	ArgGluIleLeuLysGluProValHisGlyValTyrTrpAspProSerLysAspLeuIle	480
Qy	1651	GCCGAGATCCAGAGAGGCGCCAGCACCGATGGACCTACCATGATCTTACAGGAGCCCTTC	1710

```
Db 481 AlaGluileGlnLysGlnGlyHieGlyGlnTrpThrTyrglnileTyrglnGluProPhe 500
Qy 1711 AAGAACTGAAGACCGCAGTAGTACCCAGATGGCACCACCCACCAACGAGCTGAAG 1770
Db 501 LysAsnLeuLysThrGlyLysTyraAlaArgMetArgGlyAlaHisThrAsnAspValLys 520
Qy 1771 CAGCTGACCGGAGGCGGTGACAGAGATCGCCATCGAGAGCATCGTGATCTGGGGCAAGACC 1830
Db 521 GlnLeuAlaGluAlaValGlnArgIleSerThrGluSerIleValIleTrpGlyArgThr 540
Qy 1831 CCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTACTGG 1890
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpAlaGluTyrrp 560
Qy 1891 CAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGG 1950
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580
Qy 1951 TACCAGCTGGAGAGGAGGCCATCATCGCGCGCAGAGACCTTCTACGTGGACGCGCGGCC 2010
Db 581 TyrglnLeuGlnLysGluProIleIleGlyAlaGluThrPheTyrrValAspGlyAlaAla 600
Qy 2011 AACCGCGAGACCAAGATCGCAAGCGCGGTACGTGACCGACCGGGCGCGCAAGATC 2070
Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrrValThrAspArgGlyArgGlnLysVal 620
Qy 2071 GTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGGAGCCATCCAGCTGGCCCTG 2130
Db 621 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
Qy 2131 CAGGACAGCGCAGCAGGTGAACATCGTACCGACAGCAGGTACGCGCGCTGGGCATCATC 2190
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrrAlaLeuGlyIleIle 660
Qy 2191 CAGCCCGACCGCAGCAAGACGAGCAGCTGGTGAACAGATCATCGAGCAGCTGATC 2250
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIle 680
Qy 2251 AAGAGAGAAAGTGTACCTGAGCTGGTGGTCCCGCCACCAAGGGCATCGCGCGCAAGCAG 2310
Db 681 LysLysGlnLysValTyrrLeuAlaTrpValProAlaHisLysGlyIleGlyArgGlnGlu 700
Qy 2311 CAGATCCACAGCTGTGTGAGCAAGGCGATCCGCAAGGTGCTGTTCCTGGACGGCATCGAT 2370
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

## RESULT 13

```
US-07-743-357-8
; Sequence 8, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: RF
; US-07-743-357-8
```

## Alignment Scores:

```
Pred. No.: 1,2e-210 Length: 1003
Score: 3494.50 Matches: 651
Percent Similarity: 95.83% Conservative: 39
Best Local Similarity: 90.42% Mismatches: 27
Query Match: 76.25% Indels: 3
DB: 2 Gaps: 2
```

US-09-610-313B-30 (1-2469) x US-07-743-357-8 (1-1003)

```
Qy 220 TTCTTCGGCAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGCAGCGAGCAG 279
Db 1 PhePheArgGlnAsnLeuAlaPheProGlnGlyLysAlaArgGlnLeuSerSerGlnGln 20
Qy 280 AACCGCGCCCAACAGCCCGCAGCCAGCTGCGAGCTGCGCGC---GACAAACCCCGC 336
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSerLeu 40
Qy 337 AGCGAGCGCGCGCGCAGGCGCCAGGCGCCCTG---AAGTTCCTCCAGATCACCTG 390
Db 41 SerGluAlaGlyLysArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60
Qy 391 TGCAGCGCCCTGCTGAGCATCAAGGTGGGCGGCAGATCAAGGAGGCGCTGCTGCAC 450
Db 61 TrpGlnArgProIleValThrValLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 80
Qy 451 ACCGCGCGCCGACGACACCGCTGCTGGAGGAGATGAGCTCCCGCGCAAGTGGAGCCCAAG 510
Db 81 ThrGlyAlaAspAspThrValLeuGluMetAsnLeuProGlyLysTrpLysProLys 100
Qy 511 ATGATCGCGCGCATCGCGGCTTCATCAGGTGGCGCAGTACACCGAGATCCTGATCGAG 570
Db 101 MetIleGlyIleGlyGlyPheIleLysValArgGlnTyrrAspGlnIleLeuIleGlu 120
Qy 571 ATCTGGCGCAAGAGGCCATCGCGCAGCTGATCGGCGCCACCCCGCTGAGAACATCATC 630
Db 121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140
Qy 631 GGCGCGCAACATGTGACCCAGCTGGGCTGCACCTGCACTTCCCATCAGCCCCCATCGAG 690
Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160
Qy 691 ACCGTGCGCGTGAAGTGAAGCCCGCATGAGCGCGCCCAAGGTGAAGCATGTGCCCCCTG 750
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180
Qy 751 ACCGAGGAGAGATCAAGGCGCTGACCGCCATCTCGCGAGGAGATGGAGAGAGGGCAAG 810
Db 181 ThrGluGlnLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLys 200
```

QY 811 ATCCAGATCGCCCGGAGAACCCCTACACACCCCGCTGCTGCATCAAGAGAG 870  
Db : : : : :  
QY 201 IleserlylileGlyProGluuAsnProTyrAsnThrProValPheAlaileLysLys 220  
Db : : : : :  
QY 871 GACAGCAACAAAGTGGCGAAGCTGGTACTTCGCGAGCTCAACAAAGCAGCCAGGAC 930  
Db : : : : :  
QY 221 AspSerThrLysTyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240  
QY 931 TTCTGGAGGTGAGCTGGGATCCCAACCCCGCGCGCTGAGAGAGAGAGCGTG 990  
Db : : : : :  
QY 241 PheTrpGluValGlnLeuGlyLysProHisProAlaGlyLeuLysLysSerVal 260  
QY 991 ACCGTGCTGGAGCTGGGAGCGCTTACTTCAAGCTGCGCTGGAGGAGCTTCGCAAG 1050  
Db : : : : :  
QY 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysGluPheArgLys 280  
QY 1051 TACACCGCTTACCATCCCGAGCATCAACACAGACCCCGCGCATCCGCTACAGTAC 1110  
Db : : : : :  
QY 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProArgIleArgTyrGlnTyr 300  
QY 1111 AACGTGCTGCCCCAGGGCTGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG 1170  
Db : : : : :  
QY 301 AsnValLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
QY 1171 ATCTGGAGCCCTTCGCGCGCGCAACCCCGAGATCGTGATCTACCATGATGACGAC 1230  
Db : : : : :  
QY 321 IleLeuGluProPheLysLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
QY 1231 CTGTACGTGGCAGCAGCTGAGATCGGCGCAGCAGCGCGCAAGATCGAGGAGCTGCGC 1290  
Db : : : : :  
QY 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgIleLysIleGluGluLeuArg 360  
QY 1291 AAGCACCTGTGCTGCGGCTTCCACACCCCGCAGAGAGCAGCAGAGAGAGAGAGAGAG 1350  
Db : : : : :  
QY 361 GluHisLeuLeuLysTyrGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
QY 1351 TTCTGTGGATGGCTAGCAGCTGACCCCGACAGTGGACCGTGGAGCCCATCGAGCTG 1410  
Db : : : : :  
QY 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTyrThrValGlnProIleValLeu 400  
QY 1411 CCCGAGAGGAGAGCTGGACCGTGAACACATCCAGAGCTGTGGCGCAAGCTGCACTGG 1470  
Db : : : : :  
QY 401 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
QY 1471 GCCAGCCAGATCTACCCCGCATCAAGGTGGCCAGCTGTGCAAGCTGCTCGCGGCGCC 1530  
Db : : : : :  
QY 421 AlaSerGlnIleTyrAlaGlyLysValLysGlnLysGlnLysLeuLeuArgGlyThr 440  
QY 1531 AAGCCCTGACCGACATGCTGCTGCTGACCGAGGAGCGAGCTGGAGCTGGCGAGAAC 1590  
Db : : : : :  
QY 441 LysAlaLeuThrGluValValGlnLeuThrLysGluAlaGluLeuGluAlaGluAsn 460  
QY 1591 CGCAGATCTTCGCGGAGCCGCTGACCGGCTGTACTACGACCCCGAGGAGCTGGTG 1650  
Db : : : : :  
QY 461 ArgGluLeuLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
QY 1651 GCCGAGATCCAGAGCAGGCGCACGACGAGGACCTACAGATCTACAGAGCCCTTC 1710  
Db : : : : :  
QY 481 AlaGluLeuGlnLysGlnGlyGlnTyrThrThrThrGlnIleTyrGlnLysProPhe 500  
QY 1711 AAGAACCTGAAGCCGCAAGTACGCCAAGATGCCACCGCCACCAACAGCAGCTGAAG 1770  
Db : : : : :  
QY 501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520  
QY 1771 CAGCTGACCGAGGCGCTGCAGAAATGCCATGAGAGCATCGTGATCTGGGGCAAGACC 1830  
Db : : : : :  
QY 521 GlnLeuThrGluAlaValGlnLysValAlaThrGluSerIleValIleTyrGlyLysThr 540  
QY 1831 CCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGCTGGACCGACTACTGG 1890  
Db : : : : :  
QY 541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560  
QY 1891 CAGGCCACCTGGATCCCGCGAGTGGAGTTCTGTGAACACCCCGCTGCTGGAGCTGTGG 1950

Db 561 GlnAlaThrTrpIlePheProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
QY 1951 TACCAGCTGGAGAGAGGAGCCCATCATCGCGCCCGAGACCTTCTAGCTGGAGCGCGCCGCC 2010  
Db : : : : :  
QY 581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 2011 AACCGCAGAGCAAGATCGGCAAGCGCGCTACGTGACCGAGCCCGGGCGCGAGAGATC 2070  
Db : : : : :  
QY 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
QY 2071 GTGAGCTGACCGCAGAGACCAACAGAGACCGAGCTGCGAGGCGCATCCAGCTGGCCCTG 2130  
Db : : : : :  
QY 621 ValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
QY 2131 CAGCAGACCGCAGCAGGAGTGAAATCGTGACCGCAGCAGCAGCTGCGCTGGCATCATC 2190  
Db : : : : :  
QY 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
QY 2191 CAGCGCCAGCCGCAAGAGCAGGAGGAGCTGGTGAACAGCATCATCGAGCAGCTGATC 2250  
Db : : : : :  
QY 661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680  
QY 2251 AAGAAGAGAGAGGTGTACTGAGCTGGGTGCGCCGCCACAGGCGCATCGCGGCAACGAG 2310  
Db : : : : :  
QY 681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsnGlu 700  
QY 2311 CAGATCCACAGAGCTGGTGACAGAGGCGCATCGCAAGGTGCTGTCTCGGAGCGGATCGAT 2370  
Db : : : : :  
QY 701 GlnValAspArgLeuValSerThrGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

## RESULT 14

US-07-743-357-7  
; Sequence 7, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/07743,357  
; APPLICATION NUMBER: 21-AUG-1991  
; FILING DATE: 424  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1004 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein



```
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: SF2
; US-07-743-357-7

Alignment Scores:
Pred. No.: 3,52e-210 Length: 1004
Score: 3487.00 Matches: 650
Percent Similarity: 95.42% Conservative: 38
Best Local Similarity: 90.15% Mismatches: 29
Query Match: 76.09% Indels: 4
DB: 2 Gaps: 2

US-09-610-313B-30 (1-2469) x US-07-743-357-7 (1-1004)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCTCCCGAGGAGCGCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACGGGCGCAACAGCCGCCACAGCGCGGAGCTCGAGTGGCGCGC-----GACAACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCCAGCGCACCTCG-----AACTTCCCGCCAGATCACCC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCTCTGTGTGAGCATCAAGGTGGCGGCGCGAGATCAAGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGGCGCGAGCACCGTGTCTGGAGAGATGAGCTGCGCGGCAAGTGGAGGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGGAGTACGACGACGATTCCTGATC 567
Db 101 LysMetIleGlyIleGlyIleGlyPheIleLysValArgGlnIleThrValProVal 120
QY 568 GAGATCTGCGGCAAGAGCGCATCGGCACCGTGTGATCGGCGCCCGCCACCGCCGCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATGCGCGCGCAATGCTGACCGAGCTGGGTGACCTGACCTGAACTTCCCGATCAGCCCGCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGGCGTGAAGCTGAAGCGCGCGCATGACGCGCGCCCAAGGTGAAGCAGTGGSCCC 747
Db 161 GluThrValProValLysLeuLysProGlyWecAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGGAGAGAGATCAAGGCGCCTCAGCGCCCATCTGCGAGGAGATGAGAGAGGCGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATACCAAGATTCGCGCCCGCGAGAACCCCTACAAACACCGCCCGTGTTCGCCCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTrpAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCGAGCTGGTGGACTTCCGCGAGCTTGAACAAGCGCCAGCCAG 927
Db 221 LysAspSerThrLysTrpLysLeuValAspPheArgGluLeuAsnLysAspGln 240
QY 928 GACTTCTGGGAGGTGACAGCTGGGCGATCCCCACCGCCCGCGCTGAAAGAGAGAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProIleProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGACGTGGGCGAGCGCTTCTAGCGTGGCCCTTGGACGAGGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaThrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCCCTTACCATCCCGAGCATCAACAACAGAGACCGCCCGGCTCCCTACCAG 1107
Db 281 LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln 300
QY 1108 TACAACGTGCTGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCCTGGAGCCCTTCCGCGCCCAACCCGAGATCGTGATCTACAGATACATGGAC 1227
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyGlnTyMetAsp 340
QY 1228 GACCTGTAGTGGGCGAGCAGCTGGAGATCGGCCACCGCGCCCAAGATCGAGGAGCTG 1287
Db 341 AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 360
QY 1288 CGCAACGACCTGTGCTGGGCTTCAACCCCGCACCAAGAGACCAAGAGAGGCC 1347
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1348 CCCTTCTGTGGATGGCTACGAGCTGCACCCGCAAGTGGACCGCTGCAGCCCATCGAG 1407
Db 381 ProPheLeuTrpMetGlyTyGluLeuHisProAspLysTrpThrValGlnProIleMet 400
QY 1408 CTGCCCGAGAGGAGAGCTGACCCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAAC 1467
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1468 TGGGCGAGCGAGATCTATCCCGCGCATCAAGTGGCGGCGAGCTGTGCAAGCTGTCCGGCG 1527
Db 421 TrpAlaSerGlnIleTyAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
QY 1528 GCCAAGCGCTGACCGACATCGTGGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAG 1587
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1588 AACCGGAGATCTCTGCGCGAGCCCGTGCACGCGGTGTACTACACCCCGAGCAGGACCTG 1647
Db 461 AsnArgGluIleLeuLysGluProValHisGluValTyThrAspProSerLysAspLeu 480
QY 1648 GTGGCGAGATCCAGAAAGCGGCGCGAGCGGCGAGTGGACCTTACAGATCTACAGGAGCCC 1707
Db 481 ValAlaGluIleGlnLysGlnGlyGlnTyThrTrpThrTyGlnIleTyGlnGluPro 500
QY 1708 TTCAGAAGCTGAAGACCGGCAAGTACGCCAAGTACGCCACCGCCCGCACCAACGAGCTG 1767
Db 501 PheLysAsnLeuLysThrGlyLysTyAlaArgMetArgGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGGCGCGTGCAGAAGATCGCCATCGAGAGCATCGTGTCTGGGGCAAG 1827
Db 521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540
QY 1828 ACCCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGGAGACCTGGCGACTAC 1887
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTy 560
QY 1888 TGGCAGCGCACCTGGATCCCGGAGTGGAGTTCGTGAACACCGCCCGCCCTGGTGAAGCTG 1947
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1948 TGGTACAGCTGGAAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCC 2007
Db 581 TrpTyGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyValAspGlyAla 600
QY 2008 GCCAACCGGAGACCAAGATCGGCAAGCGCGGTAGTGACCGACCGGCGCGCGAGAG 2067
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyValThrAspArgGlyArgGlnLys 620
QY 2068 ATCGTGTAGCTGACCGGAGACCAACCAAGACGAGCTGCGAGGCGCATCTTACGTGGAGCTGGCC 2127
Db 621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2128 CTGCAGGACAGCGCGCAGGAGTGAACATCGTGAACCGGAGCAGGCTACGCGCTGGGCATC 2187
```

```
Db 641 LeuGlnAaspSergGlyLeuGluValAsnIleValThrAaspSergGlnTyrAlaLeuGlyIle 660
Qy 2188 ATCCAGGCCAGCCGACCAAGAGCAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTG 2247
Db 661 IleGlnAlaGlnProAaspLysSerGlnSerGluLeuValSerGlnIleIleGluGlnLeu 680
Qy 2248 ATCAAGAAGGAGAGGTGTACTGAGCTGGTGGTGGCCGCCCAAGAGGATCGGGGGCAAC 2307
Db 681 IleIysIysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
Qy 2308 GAGCAGATCGACAGCTGGTGGCAAGGCGCATCCGCAAGGTGCTTCTCGAGCGGCATC 2367
Db 701 GluGlnValAaspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720
Qy 2368 GAT 2370
Db 721 Asp 721

RESULT 15
US-09-319-588C-6
; Sequence 6, Application US/09319588C
; Patent No. 6509018
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAULIERE, Philippe
; APPLICANT: LOUSSERT-AJAXA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSI, Francoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/09/319,588C
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-319-588C-6

Alignment Scores:
Pred. No.: 1,43e-193 Length: 1014
Score: 3221.50 Matches: 587
Percent Similarity: 90.80% Conservative: 74
Best Local Similarity: 80.63% Mismatches: 56
Query Match: 70.29% Indels: 11
DB: 4 Gaps: 3

US-09-610-313B-30 (1-2469) x US-09-319-588C-6 (1-1014)
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGCGAGGCCCGCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProAaspAsn 20
Qy 280 AAC-----CGCGCAACAGCCCAACAGCCGAGCTGCAGGTG----- 318
Db 21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40
Qy 319 ---CGCGGCAACACCCCGAGGCGCCCGCGCGCGAGCCAG-----GGCACC 366
Db 41 HisThrGlyGluGlyAspAlaGlyLeuProGlyLeuAspArgGluLeuSerValProthr 60
Qy 367 CTGAACCTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGGC 426
Db 61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
Qy 427 CAGATCAAGGAGGCGCTCTGCGACCGCGCGCGAGCACCGCTGCTGGAGGAGATGAGC 486
```

```
Db 81 GluValArgGluAlaLeuLeuAaspThrGlyAlaAaspAaspThrValIleGluGluLeuGln 100
Qy 487 CTGCGCGGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGGCG 546
Db 101 LeuGluGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArg 120
Qy 547 CAGTACGACCAAGATCTCGATCGAGATCTGCGGCAAGAGGCGCATCGGACCGTGTGATC 606
Db 121 GlnTyrAaspAsnIleThrValAaspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140
Qy 607 GGCCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGGACCCCTG 666
Db 141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160
Qy 667 AACTTCCCATCAGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGATGAGCGCGC 726
Db 161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetaspGly 180
Qy 727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGC 786
Db 181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuAargGluIleCys 200
Qy 787 GAGGAGATGAGAGAGGAGGCGCAAGATCACCAAGATCGCGCGGAGAACCCCTACACACC 846
Db 201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220
Qy 847 CCCGTGTTGGCATCAAGAAGAAGACAGCAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db 221 ProllePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240
Qy 907 GAGTGAACAGCGACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCCACCCCGCC 966
Db 241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260
Qy 967 GGCTGGAAGAAGAGAGAGCGTGCACCGTGTCTGACCTGGGCGAGCGCTACTCTACGCTG 1026
Db 261 GlyLeuLysGlnLysLysSerValThrValLeuAaspValGlyAspAlaTyrPheSerCys 280
Qy 1027 CCCCTGACGAGGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACACGAG 1086
Db 281 ProlleAspLysAspPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGlu 300
Qy 1087 ACCCGCGCATCCGCTACAGTACAGTGTGCTGCCCGAGGCTGGAAGGCGAGCCCGCAGC 1146
Db 301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320
Qy 1147 ATCTTCCAGACGACATGACCAAGATCCTGAGCGCTTCCCGCGCCGCAACCCCGAGATC 1206
Db 321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340
Qy 1207 GTGATCTACCAAGTACATGGACGACCTGTACCTGGCGCAGCGACCTGGAGATCGGCCAGC 1266
Db 341 IleIleTyrGlnTyrMetAspAspLeuTrpValGlySerAspLeuGluLeuAlaGlnHis 360
Qy 1267 CGCGCCCAAGATCGAGAGCTGCGCAAGCACCTGTGCTGGCTGGGCTTCCACACCCCGCAG 1326
Db 361 ArgGluAlaValGluAaspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380
Qy 1327 AAGAAGCACCAAGAAGGAGCCCGCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAGAAC 1386
Db 381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400
Qy 1387 TGGACCGTGGACGCCCATCGAGCTGCCGAGAGGAGAGCTGACCGTGAACGACATCCAG 1446
Db 401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420
Qy 1447 AAGCTGTGGGCAAGCTGAACCTGGGCGCCAGCAGATCTACCCCGGCGATCAAGGTGGCGC 1506
Db 421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440
Qy 1507 CTGTGCAAGCTGTGCGCGCGCCCAAGCCCTGACCGGACATCGTGGCTGCGCTGACCGGAG 1566
```

```

Db      441  LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValaenPheThrGluGlu  460
QY      1567  GCGAGCTGAGCTGGCGGAGAACCGGAGATCTCGCGGAGCCCGTGCACGGCGTGAC  1626
Db      461  AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr  480
QY      1627  TAGCACCCACGACGAGGACCTGGTGGCGGAGATCCAGAAGCAGGCGCCACGACACGATGGACC  1686
Db      481  TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnTyrThr  500
QY      1687  TACAGATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC  1746
Db      501  TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg  520
QY      1747  ACCGCCACACGACGAGCTGAGAGCTGACCGAGCGCGTGCAGAGATCGCCATCGAG  1806
Db      521  SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu  540
QY      1807  AGCATCGTGATCTGGGCAAGACCCCAAGTTCGCTCGCCATCCAGAAGGAGACCTGG  1866
Db      541  SerIleValIleTyrGlyLysThrProLysPheArgLeuProValGlnLysGluValTyr  560
QY      1867  GAGACCTGTGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC  1926
Db      561  GluAlaTyrThrAspHisTyrGlnAlaThrTyrIleProGluTyrGluPheValAsn  580
QY      1927  ACCCCCCCTGTGTGAGCTGTGGTACGACCTGAGAGAGAGGCCCATCATCGGGCGGAG  1986
Db      581  ThrProProLeuValLysLeuTyrTyrGlnLeuGluThrGluProIleSerGlyAlaGlu  600
QY      1987  ACCTTCTACGTGGACGGCGCGCAACCGGAGACCAAGATCGCAAGCGCGCTACGTG  2046
Db      601  ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal  620
QY      2047  ACCGACCGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG  2106
Db      621  ThrAspArgGlyArgGlnLysValIleSerIleAlaAspThrThrAsnGlnLysAlaGlu  640
QY      2107  CTGAGGCCATCCAGCTGGCGCTGCAGGACAGCGGCGAGCGGTGAACATCGTGACCGAC  2166
Db      641  LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp  660
QY      2167  AGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCGACAGAGCGAGAGCGAGCTGGTG  2226
Db      661  SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal  680
QY      2227  AACGAGATCATCGAGAGCTGATCAAGAGGAGAGGTGTACTGTAGCTGGTGCCCGCC  2286
Db      681  SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla  700
QY      2287  CACAAGGGCATCGCGGCAAGCAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG  2346
Db      701  HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys  720
QY      2347  GTGCTGTCTCTGGACGGCATCGAT  2370
Db      721  IleLeuPheLeuAspGlyIleGlu  728

```

Search completed: June 2, 2005, 05:11:22  
Job time : 112.083 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 04:23:54 ; Search time 153.54 Seconds

(without alignments)

11117.427 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 4583

Sequence: 1 gtcagccaccatgcccga.....gggctagcaccggtgaattc 2469

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters: 2931222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2\_1/USPRO spooll/US09610313/runat\_31052005\_155138\_15193/app\_query.fasta\_1.7893  
-DB=Published Applications AA -QFMT=fastan SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09610313 @CGN 1 1 550 @runat 31052005\_155138\_15193  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\* Query

Result

No.	Score	Match	Length	DB	ID	Description
1	3608	78.7	998	16	US-10-332-413-4	Sequence 4, Appli
2	3539.5	77.2	850	16	US-10-332-413-10	Sequence 10, Appl
3	3535	77.1	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3513	76.7	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3485	76.0	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3485	76.0	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3468	75.7	1003	16	US-10-325-468-23	Sequence 23, Appl
8	3466	75.6	1003	16	US-10-325-468-35	Sequence 35, Appl
9	3465	75.6	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3465	75.6	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3457	75.4	995	15	US-10-296-734-2	Sequence 2, Appli
12	3443	75.1	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3289.5	71.8	1006	15	US-10-296-734-1471	Sequence 1471, Ap
14	3221.5	70.3	1014	14	US-10-301-661A-6	Sequence 6, Appli
15	3135.5	68.4	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3135.5	68.4	1350	16	US-10-380-641-35	Sequence 35, Appl
17	3135.5	68.4	1350	17	US-10-636-730-35	Sequence 35, Appl
18	2860.5	62.4	875	10	US-09-952-060-6	Sequence 6, Appli
19	2860.5	62.4	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2860.5	62.4	875	16	US-10-380-641-6	Sequence 6, Appli
21	2860.5	62.4	875	17	US-10-636-730-6	Sequence 6, Appli
22	2855	62.3	850	10	US-09-952-060-2	Sequence 2, Appli
23	2855	62.3	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2855	62.3	850	16	US-10-380-641-2	Sequence 2, Appli
25	2855	62.3	850	17	US-10-636-730-2	Sequence 2, Appli
26	2827	61.7	562	9	US-09-735-487-14	Sequence 14, Appl
27	2827	61.7	562	14	US-10-342-188-14	Sequence 14, Appl
28	2823	61.6	560	15	US-10-399-920-2	Sequence 2, Appli
29	2821	61.6	561	14	US-10-059-271-87	Sequence 87, Appl
30	2815	61.4	562	14	US-10-102-622-10	Sequence 10, Appl
31	2814.5	61.4	875	10	US-09-952-060-8	Sequence 8, Appli
32	2814.5	61.4	875	15	US-10-168-217A-8	Sequence 8, Appli
33	2814.5	61.4	875	17	US-10-380-641-8	Sequence 8, Appli
34	2814.5	61.4	875	17	US-10-636-730-8	Sequence 8, Appli
35	2809	61.3	850	10	US-09-952-060-4	Sequence 4, Appli
36	2809	61.3	850	15	US-10-168-217A-4	Sequence 4, Appli
37	2809	61.3	850	16	US-10-380-641-4	Sequence 4, Appli
38	2809	61.3	850	17	US-10-636-730-4	Sequence 4, Appli
39	2796	61.0	560	14	US-10-205-641-1	Sequence 1, Appli
40	2786.5	60.8	561	14	US-10-102-622-12	Sequence 12, Appl
41	2771	60.5	560	11	US-09-725-652-1	Sequence 1, Appli
42	2740	59.8	546	8	US-08-808-031A-29	Sequence 29, Appl
43	2727	59.5	43	979	US-10-271-181B-118	Sequence 118, App
44	2626	57.3	851	14	US-10-059-271-88	Sequence 88, Appl
45	2609	56.9	999	15	US-10-346-000A-3	Sequence 3, Appli

# ALIGNMENTS

## RESULT 1

US-10-332-413-4  
; Sequence 4, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swanstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdool  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924

; PRIOR FILING DATE: 2000-09-15									
; NUMBER OF SEQ ID NOS: 32									
; SOFTWARE: PatentIn version 3.1									
; SEQ ID NO 4									
; LENGTH: 998									
; TYPE: PRT									
; ORGANISM: Human immunodeficiency virus type 1									
US-10-332-413-4									
Alignment Scores:									
Pred. No.: 3.5e-168 Length: 998									
Score: 3608.00 Matches: 670									
Percent Similarity: 97.63% Conservative: 29									
Best Local Similarity: 93.58% Mismatches: 17									
Query Match: 78.73% Indels: 0									
DB: 16 Gaps: 0									
US-09-610-313B-30 (1-2469) x US-10-332-413-4 (1-998)									
QY	223	TTCCGGAGGACCTGGCTTCCCGCAGGCAAGGCCGCGAGTTCCTCCAGCAGCAGAAC	282	Db	261	ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe	280		
Db	1	PheArgGluAsnLeuAlaPheProGlnGlyGluAlaAaArgGluPheProSerGluGlnThr	20	Db	281	ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro	300		
QY	283	CGGCGCAACAGCCCCACGACCGCGAGCTGCAGTGGCGGGCGGCAACCCCGCGACGAG	342	QY	1123	CAGGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGGCC	1182		
Db	21	ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu	40	Db	301	GlnGlyTyrPheGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro	320		
QY	343	GCCGGCGCGAGCGCCAGCGGCACCTTGAATCTCCCGCAGATCACCTGTGGCAGCGCCC	402	QY	1183	TTCCGGCGCCGCAACCCCGAGATCGTGATCTACAGTACATGACGACCTGTAGTGGC	1242		
Db	41	ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro	60	Db	321	PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly	340		
QY	403	CTGGTGAAGCATCAAGTGGCGGCCAGATCAAGAGGCCCTCTCGACACCGCGCGCGAC	462	QY	1243	AGCGACTGAGATCGCGCAGCACCAGCGCAGATCGAGAGCTGGCAGACCTGCTG	1302		
Db	61	LeuValSerIleLysIleGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp	80	Db	341	SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuArgGluHisLeuLeu	360		
QY	463	GACACCGTGTGGAGGAGATGAGCTGCCCGCAAGTGGAGCCCAAGATGATCGCGGC	522	QY	1303	CGCTGGGCTTCCACCCCGACCAAGAGCACCAGAGGAGCCCTTCTGTGTGATG	1362		
Db	81	AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly	100	Db	361	LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMet	380		
QY	523	ATCGCGGCTTCATCAAGTGCGCCAGTACGACGATCTGATCGAGATCTCGCGCAAG	582	QY	1363	GGCTAGAGCTGCACCCGACAAAGTGGACCGTCCAGCCCATCGAGTGCCTGGAGAGG	1422		
Db	101	IleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLys	120	Db	381	GlyTyrGluLeuHisProAspLysTyrThrValGlnProIleGlnLeuProGluLysAsp	400		
QY	583	AAGGCCATCGGACCGCTGTGATCGGCCCGCCACCCCGTGAACATCATCGCGCCGACATG	642	QY	1423	AGCTGACCGTGAACCATCCAGAGCTGGTGGCAAGCTGAACCTGGCGCCAGCCAGATC	1482		
Db	121	LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleLeuGlyArgAsnMet	140	Db	401	SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle	420		
QY	643	CTGACCCAGCTGGGCTGCACCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGTG	702	QY	1483	TACCCCGGATCAAGTGGCCAGCTGTGCAAGCTGTGGCGGGCGCAAGGCCCTGACC	1542		
Db	141	LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal	160	Db	421	TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr	440		
QY	703	AAGCTGAAGCCCGCATGAGCGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAG	762	QY	1543	GACATCGTCCCTGACCGAGAGCGCGAGCTGGAGTGGCCAGAACCCCGAGATCCAG	1602		
Db	161	LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluLys	180	Db	441	AspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluLeuLeu	460		
QY	763	ATCAAGCCCTGACCGCCATCTCGGAGGAGATGGAGAAGGCGCAGATCACCAAGATC	822	QY	1603	CGCGAGCCGTGCACGGCGTGTACTACGACCCCGCAGCAGGACCTGGTGGCGAGATCCAG	1662		
Db	181	IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle	200	Db	461	LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln	480		
QY	823	GGCCCCGAGAACCCCTACAAACACCCCGTGTTCGCCCATCAAGAAGAAGGACACCAAG	882	QY	1663	AAGCAGGCGCCACGACGATGGACCTTACCAGATCTTACCAGAGGCCCTTCAAGAACCTGAAG	1722		
Db	201	GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys	220	Db	481	LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnLysProPheLysAsnLeuLys	500		
QY	883	TGGCGCAGCTGGTGACTTCGCGAGCTGAACACGCGCCAGCCAGACTTCTGGGAGGTG	942	QY	1723	ACCGCAAGTACGCCAAGATGCGCAGCGCCACCAACACGACGTGAAGAGCTGACCGAG	1782		
Db	221	TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal	240	Db	501	ThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThrGlu	520		
QY	943	CAGCTGGGCATCCCCACCCCGCGCCCTGAAGAAGAAGAGCGTGTACCTGTCTGAC	1002	QY	1783	GCGTGCAGAAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAGACCCCGAGTTCGCG	1842		
Db	241	GlnLeuGlyIleProHisProAlaGlyLeuAlaGlyLeuLysLysLysSerValThrValLeuAsp	260	Db	521	AlaValGlnLysIleSerLeuGluSerIleValIleTyrPheLysThrProLysPheArg	540		
QY	1003	GTGGGCGACCTTACTGACGTGCCCTTGGACGAGGACTTCCGCAAGTACACCGCCTTC	1062	QY	1843	CTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGCTACTGTGACGCGCCACCTGG	1902		
Db				Db	541	LeuProIleGlnLysGluThrTrpGluIleTrpThrAspTyrTrpGlnAlaThrTrp	560		
QY				QY	1903	ATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGTACCGAGCTGGAG	1962		
Db				Db	561	IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTyrTrpGlnLeuGlu	580		
QY				QY	1963	AAGAGCCCATCATCGCGCCGAGAGCTTCTACGTGGAGCGCGCGCCCAAGCCGAGACCC	2022		
Db				Db	581	LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr	600		
QY				QY	2023	AAGATCGCAAGCCCGCTACGTGACCGACCGCGCGCCCGCAGAGATCGTGAGCCTGACC	2082		
Db				Db	601	LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer	620		
QY				QY	2083	GAGACCAACCAAGAGACCGAGCTGACAGCCATCAGCTGGCCCTGACGAGCAGCGGC	2142		
Db				Db	621	GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu	640		

```
QY 2143 ACCGAGGTGAACATCTGTACCGACAGGCGAGTACGCTTGGCGCATCATCCAGGCCCGCAGCCC 2202
Db 641 SerGluValAsnIleThrAspSerGlnTyrAlaLeuGlyIleGlnAlaGlnPro 660
QY 2203 GACAAAGCGAGAGCGAGCTGTGTAAACAGATCATCGAGCGCTCATCAAGAGGAGAAG 2262
Db 661 AspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluArg 680
QY 2263 CTGTACTGTAGCTGGGTCCCGCCACCAAGGGCATCGCGCGCAACGACGACGATCGACAAG 2322
Db 681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700
QY 2323 CTGTGACCAAGGCGATCCGAGGTGCTTCTTCTGTGAGCGGCATCGAT 2370
Db 701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716

RESULT 2
US-10-332-413-10
; Sequence 10, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
; APPLICANT: Johnstone, Robert Edward
; APPLICANT: Swanson, Ronald Ivar
; APPLICANT: Morris, Lynn
; APPLICANT: Karim, Salim Abdool
; APPLICANT: Williamson, Carolyn
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H
; TITLE OF INVENTION: Subtype Isolates, Their Genes and Modifications and Derivatives T
; FILE REFERENCE: 45669-281993
; CURRENT APPLICATION NUMBER: US/10/332,413
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: PCT/IB01/01208
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216,995
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: ZA 2000/3437
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: ZA 2000/4924
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 850
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-10

Alignment Scores:
Pred. No.: 7,62e-165 Length: 850
Score: 3539.50 Matches: 669
Percent Similarity: 91.18% Conservative: 34
Best Local Similarity: 86.77% Mismatches: 45
Query Match: 77.23% Indels: 23
DB: 16 Gaps: 4

US-09-610-313B-30 (1-2469) x US-10-332-413-10 (1-850)
QY 97 TGCTTCACTGCGCAGAGGAGGCGCACATCGCCGCGCACTGCGCGCGCCCGCGCAAG 156
Db 84 CysAlaThrValGlyLys-----GlyAspArgCysGlyProLeuArgTyrTyr 99
QY 157 GGCTGTCTGGAAG-----TGCGCAGAGGAGGCGCACCATG 192
Db 100 AlaserTrpArgLysGlyAspValLeuGlnGlyAspValGlyArgGlnGlyPheProSer 119
QY 193 AAGAGTGCACCGAGCGCCAG-----GCCAACTTCTTCCCGAGGACCTG--- 237
Db 120 HisAspValValLysArgProValProValProSerLeuHisAlaCysArgSerThrLeuGlu 139
QY 238 -----GCTTCCCCCGAGGCGCGCGCGAGTTCCTCCCGAGGACGAG 279
Db 140 AspProArgValProSerSerPheProGlnGlyProAlaArgGlnPheProSerGluGln 159
```

```
QY 280 AACCGCGCCAAACAGCCCCACAGCGCGCGAGCTGCGCGCGGCGACAAACCCCGCAGC 339
Db 160 ThrArgAlaAsnSerProThrSerArgGluLeuValArgArgAspAsnProArgSer 179
QY 340 GAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 399
Db 180 GluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 199
QY 400 CCCCTGGTGTAGCATCAAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459
Db 200 ProLeuValSerIleLysIleGlyGlnThrArgGluAlaLeuLeuAspThrGlyAla 219
QY 460 GACGACACCGTGTGTGAGGAGATGAGCTGCCCGCGCAAGTGGAAAGCCCAAGATGATCGGC 519
Db 220 AspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGly 239
QY 520 GGCATCGCGCGCTTCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579
Db 240 GlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 259
QY 580 AAGAAAGCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639
Db 260 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleLysArgAsn 279
QY 640 ATGCTGACCCAGCTGGGCTGCACCTTGAACCTTCCCATTCAGCCCGCGCGCGCGCGCG 699
Db 280 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 299
QY 700 GTGAAGCTGAAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 759
Db 300 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluVal 319
QY 760 AAGATCAAGCCCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 819
Db 320 LysIleLysAlaLeuThrAlaIleCysGluGlnMetGluLysGluGlyLysIleThrLys 339
QY 820 ATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 879
Db 340 IleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysGluAspSerThr 359
QY 880 AAGTGGCGCGAGCTGTGTGAGCTTCCGCGAGCTGAACAGCGCGCGCGCGCGCGCGCG 939
Db 360 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 379
QY 940 GTGCGAGTGGCGCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 999
Db 380 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu 399
QY 1000 GAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1059
Db 400 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAla 419
QY 1060 TTCACCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
Db 420 PheThrIleProSerIleAsnAsnGlnThrProGlyIleArgTyrGlnTyrAsnValLeu 439
QY 1120 CCCGAGGCTGGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1179
Db 440 ProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSerMetThrLysIleLeuGlu 459
QY 1180 CCCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1239
Db 460 ProPheArgAlaLysAsnProGluIleValIleTyrGlnTyrMetAlaIleLeuTyrVal 479
QY 1240 GGCAGCGAGCTGTGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1299
Db 480 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 499
QY 1300 CTGGCTGGGCGCTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1359
Db 500 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 519
```



```
QY 1360 ATGGGCTACGAGTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAG 1419
Db 520 MeGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLys 539
QY 1420 GAGAGCTGGACCGTGAACACATCCAGAACCTGGTGGCAAGCTGAACCTGGCGACCCAG 1479
Db 540 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpThrSerGln 559
QY 1480 ATCTACCCCGCATCAAGTGGCGCAGCTGCAAGCTGCTGCGCGCGCCAGGCCCTG 1539
Db 560 IleTyrProGlyLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 579
QY 1540 ACCGACATCTGCCCTGCACCGAGGAGCGCGAGCTGGAGCTGCCGAGAACCCGAGATC 1599
Db 580 ThrAspIleValProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIle 599
QY 1600 CTGCGGAGCCCGTGCACCGCGTGTACTACGACCCCGCAAGCACTGTGTGGCGAGATC 1659
Db 600 LeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIle 619
QY 1660 CAGAGCAGGCGCCACGACGAGTGGACCTACAGATCTACGAGGACCTTCAAGAACCTG 1719
Db 620 GlnLysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 639
QY 1720 AAGACCGGCAAGTACGCAAGATGCGCACCGCCCGCACACACGACGTAAGCAGCTGACC 1779
Db 640 LysThrGlyLysTyrAlaLysArgArgThrHisThrAsnAspValLysGlnLeuThr 659
QY 1780 GAGGCGGTGCAGAAATCGCCATGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTC 1839
Db 660 GluAlaValGlnLysIleSerLeuGluSerIleValThrTrpGlyLysTrpProLysPhe 679
QY 1840 CGCTGCCCATCAGAAAGGAGACCTGGGAGACCTGTGGACCCACTCTGGAGGCCACC 1899
Db 680 ArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThr 699
QY 1900 TGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGGTACCAGCTG 1959
Db 700 TrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrpTyrGlnLeu 719
QY 1960 GAGAAGAGCCCATCATCGCGCGGAGACCTTCTACGTGACCGCGCGCCGCAACCGGAG 2019
Db 720 GluLysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 739
QY 2020 ACCAAGATCGCAAGCGCGCTAGTGCACCGCGCGCGCGAGAGATCTGACGCTG 2079
Db 740 ThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeu 759
QY 2080 ACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCGCTGCAGGACAGC 2139
Db 760 SerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 779
QY 2140 GGCAGCGAGTGAACATCGTACCGACGACGAGTACGCGCTGGCGCATCATCGAGGCCAG 2199
Db 780 GluSerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 799
QY 2200 CCGGACAAAGACGAGCAGCTGCTGAACACAGATCATCAGACGAGCTGATCAAGAGGAG 2259
Db 800 ProAspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 819
QY 2260 AAGGTGTACCTGAGCTGGGTGCCCGCCACCAAGGGCATCGCGCGCAACGAGCAGATCGAC 2319
Db 820 ArgAlaTyrLeuSerTrpValProAlaHisGlyIleGlyGlyAspGluGlnValAsp 839
QY 2320 AAGTGTGTGACGAGGCGCATCCCGAAGTGTCTG 2352
Db 840 LysLeuValSerSerGlyIleArgLysValLeu 850
```

## RESULT 3

```
US-10-283-847-17
; Sequence 17, Application US/10283847
; Publication No. US20030162720A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Nicolette, Charles A.
; APPLICANT: Walker, Bruce
; TITLE OF INVENTION: THERAPEUTIC ANTI-HIV (IV9) COMPOUNDS
; FILE REFERENCE: GZ 2112.00
; CURRENT APPLICATION NUMBER: US/10/283,847
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 60/345,116
; PRIOR FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus
US-10-283-847-17
```

```
Alignment Scores:
Pred. No.: 1,28e-164 Length: 1003
Score: 3535.00 Matches: 659
Percent Similarity: 96.12% Conservative: 34
Best Local Similarity: 91.40% Mismatches: 24
Query Match: 77.13% Indels: 4
DB: 14 Gaps: 2
```

US-09-610-313B-30 (1-2469) x US-10-283-847-17 (1-1003)

```
QY 220 TTCTTCGCGGAGGACCTGGCTTCCCCAGGGCAAGGCCCGCGAGTTCCCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCAACACCCCGCCAGCCCGAGCTCGAGGTG-----CGCGCGCAACACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGAGCCCGAGCCCGAGGCCACCCCTG-----AAGTTCGCCAGATCACC 387
Db 41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
QY 388 CTGTGCGCGCGCGCGCTGTGAGCATCAAGGTGGCGCGCGAGATGAGGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCGACACCGCTGTGTGGAGGAGATGAGCTGCCCGCAAGTGGAGGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluWetSerLeuProGlyArgTrpLysPro 100
QY 508 AAGATATCGCGCGCATCGCGCTTCATCAAGGTGGCGCGAGTACGACGAGATCCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTGCGGCAAGAGCCATCGCGCCGCTGATCGCGCCCGCCAGCCCGCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGTGTGACCCAGCTGGGTGCACCTGCACTTCCCATCAGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGTGAAGTGAAGCCCGGATCGAGCGGCCCGCAAGGTGAAGCAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyWetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGGAGAGAGATCAAGGCCCTCAGCCCATCTCGCGAGGAGATGGAGAGAGGGGC 807
Db 181 LeuThrGluGlnLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGCGAGAACCCCTACAAACACCCCGCTGTTCGCCCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACGACCAAGTGGCGCAAGCTGGTGACTTCGCGAGCTGAACAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
```

```
QY 928 GACTTCTGGAGGTGCTAGCTGGGCTATCCCTCCACCCCGCGCTTGAAGAAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGGAGCTGGGCGAGCTTACTTTCAGCGTCCCTCGAGGAGCTTCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArg 280
QY 1048 AAGTACACCGCTTTCACCATCCCGAGCATCAACACGAGACCCCGGCGATCCGCTACCAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGCTGCCCGAGGCTGAGGGGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCTCTGGAGCCTTCCCGCGCCGCAACCCCGAGATCGTGATCTACAGTACATGCAC 1227
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY 1228 GACCTGTACGTGGGCGAGCTGAGATCGGCGAGCACCCGCGCAAGATCGAGGAGCTG 1287
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1288 GCGAAGCACCTGCTGCGCTGGGCTTACCAACCCCGAGCAAGAGAGAGAGAGAGCC 1347
Db 361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380
QY 1348 CCCTTCTCTGGTGGGTACGAGCTGCACCCCGCAAGTGCAGCGTGCAGCCCATCGAG 1407
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1408 CTGCCCCGAGAGGAGCTGACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAC 1467
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1468 TGGGCCGACGAGATCTACCCCGCATCAAGTGCGCCAGCTGTCAAGCTCTCGCGGC 1527
Db 421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1528 GCGAAGCCCTGACGACATCGTCCCTGACCGAGGCGGCGAGCTGGAGCTGGCGGAG 1587
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
QY 1588 AACCGGAGATCTCTGGCGAGCCGTCGACCGCGTGTACTACGACCCCGAGCAAGCACTG 1647
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1648 GTGCCCGGAGATCCAGAAGCAGGCGCCACGACCATCGAGTGCACCTACAGATCTTACCAGGAGCC 1707
Db 481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
QY 1708 TTCAGAACCTGAGACCGGCAAGTACGCCAAGATGGCCACCGCCGACCAACAGCAGTG 1767
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGGCGCTGAGAGATGCCATCGCAGAGCATCGTATCTGGGGCAAG 1827
Db 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540
QY 1828 ACCCCCAAGTTCGCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACTAC 1887
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
QY 1888 TGGCAGGCGACCTCGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTCGTGAAGCTG 1947
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1948 TGGTACAGCTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTACGTGCAGCGCGCC 2007
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
```

```
QY 2008 GCCAACCCGAGAGCAAGATCGGCAAGCCCGGCTACGTGACCGACCGCGGCGGCGAGAG 2067
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2068 ATCTGTAGCTTACCGAGAGACCAACACAGAAAGACCGAGCTGCGAGCCATCCAGCTGCC 2127
Db 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
QY 2128 CTGCAAGCAGCGCGCAGGAGGTGAACATCGTGCACGACGACGACGACGACGACGACGAC 2187
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGGCGCGCGCAGAGAGCGAGCGAGCGAGCTGTGTGAACGACGACGACGACGACG 2247
Db 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGAGAGAGGTGTACTGAGCTGGGTGGCGCCGACCAAGGGGATCGGCGGCAAC 2307
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2308 GAGCAGATCCGACAAAGCTGGTGAGCAAGGCGCATCCGCAAGGTGCTGTTCTCGACGCGCATC 2367
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db 721 Asp 721

RESULT 4
US-10-634-165-9
; Sequence 9, Application US/10634165
; Publication No. US20050095581A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Nancy T
; APPLICANT: Gallo, Robert C
; APPLICANT: Wong-Staal, Flossie
; TITLE OF INVENTION: DETECTION OF HIV-1 DNA
; FILE REFERENCE: 223695
; CURRENT APPLICATION NUMBER: US/10/634,165
; CURRENT FILING DATE: 2003-08-05
; PRIOR APPLICATION NUMBER: 08/463,028
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/693,866
; PRIOR FILING DATE: 1985-01-23
; PRIOR APPLICATION NUMBER: 06/659,339
; PRIOR FILING DATE: 1984-10-10
; PRIOR APPLICATION NUMBER: 06/643,306
; PRIOR FILING DATE: 1984-08-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 9
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: T cell leukemia-lymphoma virus (HTLV)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1015)
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
US-10-634-165-9

Alignment Scores:
Pred. No.: 1.52e-163 Length: 1015
Score: 3513.00 Matches: 658
Percent Similarity: 94.27% Conservative: 33
Best Local Similarity: 89.77% Mismatches: 26
Query Match: 76.65% Indels: 16
DB: 17 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-634-165-9 (1-1015)

QY 220 TTCTTCCGCGAGGACCTGGCTTCCCGGAGGCGGCGGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
```

```
QY 268 -----CCGAGCGAGCAGAACCGCGCCCAACAGCCCAACAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGGCGCAACAACCCCGCAGCGAGCGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGGCACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGCGCCCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGGTGGCGCGCCAGATCAAGAGCGCCCTGCTGGACACCGCGCGCGCAGCACCGTG 471
Db 81 IleIysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100
QY 472 CTGGAGGAGATGACCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120
QY 532 TTTCATCAAGTGGCGCCAGTAGCAGCAGATCCTGTATCGAGATCTCGCGCAGAGGCCATC 591
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGTGACCCCTGAATCTCCCATCAGCCCATCGAGACCGTGGCCGCTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
QY 712 CCGGCGATGACGGCCCAAGGTGAAGCAGTGGCCCTGACCGCGAGGAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCGCATCTCGAGGAGATGGAGAAGGGCAAGATCAACAGATCGCGCCCGAG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
QY 832 AACCCCTACAAACCCCGCTGTTCGCCATCAAGAAGAGCAGCACCAGATGGCGCAAG 891
Db 221 AsnProIyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGGTGGACTTCCGAGCTGAACAGCGCACCCAGCATCTTCGGAGGTGCAGCTGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCCCACC CGCGCCCTGAAGAAGAAGAGCGTGACCGTGTGGACGTGGGCGCAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTTCAGCTGCGCCCTGGACGAGACTTCGCAAGTACACCGCTTCACCATCCCC 1071
Db 281 AlaIyrPheSerValProLeuAspGluAspPheArgLysIyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACACGAGACCCCGCGCATCCGCTACCACTAGCAACGTGTGCGCCCGAGGGCTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgIyrGlnIyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGGCAGCCCGAGCATCTTCCAGAGCAGCATCAACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerWetThrLysIleLeuGluProPheLysLys 340
QY 1192 CGCAACCCGAGATCGGTGATCTACCACTACATGACGACCTGTACGTGGCGAGCGACCTG 1251
Db 341 GlnAsnProAspIleValIleIyrGlnIyrMetAspAspLeuIyrValGlySerAspLeu 360
QY 1252 GAGATCGGCGACGCCCGCCAGATCGAGAGCTGGCAGCAACCTGCTGGCTGGCTGGGGC 1311
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuLeuArgTrpGly 380
```

RESULT 5

```
QY 1312 TTCACACCCCGCAGCAAGAACAGCACCAAGAGGAGCCCGCTTCTCTGTGGATGGGCTACGAG 1371
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyIyrGlu 400
QY 1372 CTGCACCCCGCAACAGTGGACCGCTGCAGCCCATCGACTCGCCCGAGAGGAGAGCTGGACC 1431
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGlnLysAspSerTrpThr 420
QY 1432 GTCAACGACATCCAGAAAGCTGGTGGGCAAGCTGAATGGGGCCAGCCAGCATCTACCCCGGC 1491
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleIyrProGly 440
QY 1492 ATCAAGGTGGCGCAGCTGTGCAGCTGCTCGCGCGGCCCAAGCCCTGACCGACATCGTG 1551
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1552 CCCTCACCAGGAGGCGCGAGCTGGAGCTGGCGCAGAACCGCGAGATCTCTCGCGGAGCCC 1611
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1612 GTGCACGGGTGTACTACGACCCCAAGGACCTGTGTGGCCGAGATCCAGAGAGCAGGGC 1671
Db 481 ValHisGlyValIyrIyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1672 CACGACAGTGGACCTTACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
Db 501 GlnGlyGlnTrpThrIyrGlnIleIyrGlnProProPheLysAsnLeuLysThrGlyLys 520
QY 1732 TAGCCCAAGATGGCGACCGCCACACCAAGCAGCTGAGCAGCTGACCGAGCGCGTGCAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1792 AAGATCGCATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCTCGCCCATC 1851
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1852 CAGAAGGAGACCTGGGAGACCTGTGCGACCGACTACTGCGAGCGCCACCTGGATCCCGCAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluIyrTrpGlnAlaThrTrpIleProGlu 580
QY 1912 TGGGAGTTCGTGAACACCCCGCTGTGTGAGCTGTGTACCGCTGGAGGAGGAGGCC 1971
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpIyrGlnLeuGluLysGluPro 600
QY 1972 ATCATCGCGCGCCGAGACCTTCTACGTGGACCGCGCGCCCAACCGCGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheIyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2032 AAGCGCGCTTACGTGACCGACCGCGCGCGCGCAGAGATCGTGAGCTGACCGACACCC 2091
Db 621 LysAlaGlyIyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2092 AACCAAGACCGCAGCTGCAGGCCATCCAGCTGGCCCTGCGAGCAGCAGCGCAGGAGTG 2151
Db 641 AsnGlnLysThrGluLeuGlnAlaIleIyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2152 AACATCGTGACCCACAGCAGTAGTACGCCCTGGGGCATCATCCAGGCCCGACCCCGCAAGAGC 2211
Db 661 AsnIleValThrAspSerGlnIyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2212 GAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTTGACTGT 2271
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValIyrLeu 700
QY 2272 AGCTGGTGGCCCGCCCAAGGCCATCGCGCGCAACGAGCAGATCGACAGCTGGTGAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2332 AAGGGCATCGCAAGTGTGTCTTCTGGACCGCATCGAT 2370
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

US-10-093-953A-26  
 ; Sequence 26, Application US/10093953A  
 ; Publication No. US20040105871A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Robinson, Harriet L.  
 ; APPLICANT: Smith, James M.  
 ; APPLICANT: Hua, Jian  
 ; APPLICANT: Moss, Bernard  
 ; APPLICANT: Amara, Rama  
 ; APPLICANT: Wyatt, Linda  
 ; APPLICANT: Earl, Patricia  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING  
 ; TITLE OF INVENTION: AN IMMUNE RESPONSE  
 ; FILE REFERENCE: 12804-005002  
 ; CURRENT APPLICATION NUMBER: US/10/093,953A  
 ; CURRENT FILING DATE: 2002-03-08  
 ; PRIOR APPLICATION NUMBER: US 60/186,364  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: US 60/251,083  
 ; PRIOR FILING DATE: 2000-12-01  
 ; PRIOR APPLICATION NUMBER: US 09/798,675  
 ; PRIOR FILING DATE: 2001-03-02  
 ; PRIOR APPLICATION NUMBER: PCT/US01/06795  
 ; PRIOR FILING DATE: 2001-03-02  
 ; PRIOR APPLICATION NUMBER: US 60/324,845  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: US 60/325,004  
 ; PRIOR FILING DATE: 2001-09-26  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 26  
 ; LENGTH: 739  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: protein encoded by construct of vaccine vector  
 ; OTHER INFORMATION: pGA2 and insert JS2 expressing clade HIV-1 VL  
 ; US-10-093-953A-26

Alignment Scores:  
 Pred. No.: 3,45e-162 Length: 739  
 Score: 3485.00 Matches: 654  
 Percent Similarity: 94.00% Conservative: 35  
 Best Local Similarity: 89.22% Mismatches: 28  
 Query Match: 76.04% Indels: 16  
 DB: 16 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-093-953A-26 (1-739)

Qy	220	TTCTTCCGCGAGGACCTGGCTTCCCGAGGCAAGGCCCGCGAGTTCCTCCGAGCGAG	279
Db	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
Qy	280	AACGCG-----GCAACAGCCGCCACCGC	303
Db	21	ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg	40
Qy	304	CGCAGCTGCGAGGTG-----CGCGCGCAACCCCGCAGCGAGCGCGCGCGCGAGCGC	357
Db	41	ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg	60
Qy	358	CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCTCGTGGAGC	411
Db	61	GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr	80
Qy	412	ATCAAGTGGCGCGCGAGATCAAGAGCCCTGTGGACACCGCGCGCGCGAGCACCGTG	471
Db	81	IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal	100
Qy	472	CTGAGGAGATGACCTGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC	531
Db	101	LeuGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly	120

Qy	532	TTCATCAAGGTGGCGCAGTACGACAGATCTGTATCGAGTCTCGGCAAGAGCCATC	591
Db	121	PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLeuAlaIle	140
Qy	592	GGCACCGTGTGATCGGCGCCACCCCGTGAACATCATCGCCGCAACATCTGACCCAG	651
Db	141	GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln	160
Qy	652	CTGGGTGACCCCTGAATCTTCCCATCAGGCCCATCGAGACCGTGGCGCTGAAG	711
Db	161	IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys	180
Qy	712	CCCGCATGACGCGCCCAAGGTGAACAGTGGGCCCTGACCGAGGAGAGATCAAGGCC	771
Db	181	ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla	200
Qy	772	CTGACCGCCATCTCGAGGAGATGGAGAAGAGGCAAGATCACCAAGATCGGCCCGCAG	831
Db	201	LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu	220
Qy	832	AACCCCTAACACACCCCGTGTTCGCCATCAAGAAGAGGACACACCAAGTGGCGCAAG	891
Db	221	AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys	240
Qy	892	CTGGTGAATCTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGC	951
Db	241	LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly	260
Qy	952	ATCCCCCACCCCGCGCCCTGAAGAAGAAGAGCGTGCACCGTGGAGCGTGGCGCAC	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp	280
Qy	1012	GCCTACTTCAAGCGTGGCGCGAGGACTTCCGCAAGTACACCGCTTCCATCCCTCC	1071
Db	281	AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACACGAGACCCCGCGCATCGGTACAGTACACAGTGCCTGCCCGCGGCTGG	1131
Db	301	SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp	320
Qy	1132	AAGGCGAGCCCGCAGCATCTTCCAGAGCAGATGACCAAGATCTCGAGGCGCTTCCGCGC	1191
Db	321	LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys	340
Qy	1192	CGCAACCCCGAGATCGTGATCTACAGTACATGACACCTGTAGTGGCGAGCGACTG	1251
Db	341	GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu	360
Qy	1252	GAGATCGCGCAGCAGCCCGCAGATCGAGGATCGGAGCTGCGCAAGACCTGCTCGCTGGC	1311
Db	361	GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly	380
Qy	1312	TTCAACACCCCGAGAGAGCAGCAGAGGAGCCCTTCTGTGTGATGGGTACGAG	1371
Db	381	LeuThrThrProAspLysLysHisGlnLysGluProPheLeuLeuTrpMetGlyTrpGlu	400
Qy	1372	CTGACCCCGCAGAGTGGACCGTGCAGCCCATCGAGTGGCGCGAGAGAGAGTGGAC	1431
Db	401	LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr	420
Qy	1432	GTGAACACATCCAGAGCTGGTGGCAAGCTGAGTGGCGGCGAGGAGTACCTACCGCGC	1491
Db	421	ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly	440
Qy	1492	ATCAAGTGGCGCGAGTGTCAAGCTGCTCGCGCGCCCAAGGCCCTGACCGACATCGTG	1551
Db	441	IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle	460
Qy	1552	CCCTGACCCGAGGAGCGGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGGAGCC	1611
Db	461	ProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluPro	480
Qy	1612	GTGCACGCGGTGTACTACGACCCCGCAGCAAGGACCTGGTGGCGGAGATCCAGAGCGGC	1671

```
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuileAlaGluIleGlnLysGlnGly 500
Qy 1672 CAGACACAGTGGACCTACAGATCTACAGAGCCCTTCAAGAACCTGAAGACGGGAAG 1731
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
Qy 1732 TAGCCAAAGTGGCAGCCGCCACCAACAGCTGAAGCAGCTGACCGAGGCCGTGCAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
Qy 1792 AAGATCCCCATGGAGACATCGTATCTGGGGGAAGACCCCAAGTTCCGCTGCCCATC 1851
Db 541 LysIleThrThrGluSerIleValIleTrpGlyThrProLysPheLysLeuProIle 560
Qy 1852 CAGAGAGACCTGGGAGACCTGTGTGACCGACTACTGTGGAGGCCACCTGGATCCCGAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
Qy 1912 TGGGAGTTCGTGAACACCCCGCTGTGTAAGCTGTGTACAGCTGGAGAGGAGCC 1971
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpLysGlnLeuGluLysGluPro 600
Qy 1972 ATCATCGCGCCGAGACCTTCTAGTGTGACGGCGGCCCAACCGCGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
Qy 2032 AAGCCGGCTAGTGACCGACCGCGGGCGGCGAGAGATCGTAGCCTGACCGACCAACC 2091
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
Qy 2092 AACCAAGACCGAGCTGCAGGCCATCCAGCTGCGCCCTGCAGGACAGCGCAGCGAGTG 2151
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
Qy 2152 AACATCTGTGACGACACCCAGTACGCCCTGGGCATCATCAGGCCCAAGCCCGCAAGAGC 2211
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
Qy 2212 GAGAGCAGCTGTGAACACGATCATCAGCAGCTGATCAAGAGGAGAGGTGTACCTG 2271
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLysLysLysGluLysValTyrLeu 700
Qy 2272 AGCTGGTGGCCGCCACAGGCGATCGCGGCGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlnGlnGlnGlnGlnGlnGlnGlnGln 720
Qy 2332 AAGGCGATCCGCAAGGTGCTGTTCTCTGGAGCGCATCGAT 2370
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

## RESULT 6

```
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE OF INVENTION: AN IMMUNE RESPONSE
; FILE REFERENCE: 12804-005002
; CURRENT APPLICATION NUMBER: US/10/093,953A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
```

```
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein encoded by construct of vaccine vector
; OTHER INFORMATION: pGAL and vaccine insert expressing clade B HIV-1
; OTHER INFORMATION: Gag-Pol
; US-10-093-953A-31

Alignment Scores:
Pred. No.: 3 45e-162 Length: 739
Score: 3485.00 Matches: 654
Percent Similarity: 94.00% Conservative: 35
Best Local Similarity: 89.22% Mismatches: 28
Query Match: 76.04% Indels: 16
DB: 16 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-093-953A-31 (1-739)
```

```
Qy 220 TTCTTCGGGAGACTGCGCTTCCCGAGGGGAAGCCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
Qy 280 AACCGC-----GCCAACAGCCCCACGACG 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40
Qy 304 CGCGAGCTGCAGGTG-----CGCGCGCGACAACCCCGCAGCGAGCGCGCGCGCGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
Qy 358 CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCTCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
Qy 412 ATCAAGGTGGCGCGCGCAGATCAAGGAGCGCCTCTGACACCGCGCGCGCGCGCACCGCTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
Qy 472 CTGAGGAGATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120
Qy 532 TTCATCAAGGTGGCGCGCAGTACACCGAGATCCTGATCAGATCTGCGGCAAGAGGCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
Qy 592 GGCACCGTCTGATCGSCCCCGCCCGCTGCAACATCATCGCGCGCAACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
Qy 652 CTGGGTGTCACCTGAACTTCCCGCATCGAGCCCATCGAGACCGCGTCCCGTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
Qy 712 CCGCGCATGAGCGGCCCAAGGTGAAGAGTGGCCCTGACCGAGGAGAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
Qy 772 CTGACCGCCCATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGCGCCCGAG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluLysIleSerLysIleGlyProGlu 220
```

```
QY 832 AACCCCTACACACCCCGTGTTCGCATCAAGAGAGGACAGACCAAGTGGCGAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGGAGGTGCAGTGGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCCCAACCCCGCGGCTGAAGAAGAGAGCGTGAACCGTGTGGAGCGTGGCGAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTTCAGCGTGGCCCTGGAGCGAGACTTCGCGAGTACACGGCTTCACCATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACACAGAGACCCCGGATCCGCTACAGTACAACTGCTGCTGCCCGAGGTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGGCAGCCCGAGATCTTCCAGAGCAGCATGACCAAGATCTGTGGAGCCCTTCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
QY 1192 CGCAACCCCGAGATCGTGTATCTACCGTACATGACGACCTGTAGTGGCGAGCGACCTG 1251
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360
QY 1252 GAGATCGGCCAGCAGCCGCGCAAGATCGAGAGCTGGCAAGCACTGCTGCGTGGGGC 1311
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380
QY 1312 TTCAACCCCGCCAGAGAAGCACCAGAGAGAGCCCGCTTCTGTGTGGATGGGTACGAG 1371
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGTGCCTCCGAGAGAGAGAGTGGACC 1431
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1432 GTGAACGACATCCAGAGCTGTGGGCAAGCTGAAGTGGGCGACGAGATCTACCCCGGC 1491
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440
QY 1492 ATCAAGTGTGCCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGAGATCGTG 1551
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1552 CCCGTGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCGAGATCTTCGCGGAGGCC 1611
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1612 GTGCACGCGGTACTAGCACCCAGCAAGACCTGTTGGCGGAGATCCAGAACGAGGC 1671
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1672 CACGACCAAGTGGACCTACCAATCTACCAAGAGCGCTTCAAGAACCTGAAGACCGCAAG 1731
Db 501 GlnGlyGlnThrThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1732 TACGCCAAGATGCGACCGCCCAACCAACGACGTGAAGACGTGACGAGCGCGTGGAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
QY 1792 AAGATCCCATGGAGCATCGTGTATCTGGGGCAAGACCCCAAGTTCGCGTCCCGCATC 1851
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1852 CAGAAGGAGACCTGGGAGACCTGTGTGGACCACTACTTGGCAGGCGCACCTGATCCCCGAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1912 TGGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTACCAGCTGGAGAGGAGGCC 1971
```

```
Db 581 TrpGluPheValAsnThrProLeuValLysLeuTyrTrpGlnLeuGluLysGluPro 600
QY 1972 ATCATCGCGCGCGAGACCTTCTACGTGGACCGCGCCGCAACCGCGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2032 AAGCGCGGCTACGTGACGACCGCGCGCGCGAGAGATCTGTAGCCTGACCGAGACCAACC 2091
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2092 AACCAAGAGACCGAGCTGCGAGCCATCCAGCTGCGCTTCGAGGACAGCGCGAGGAGTG 2151
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2152 AACATCTGTGACCGACAGCAGTACGCGCTGGGCGATCTCCAGGCGCCAGCCGACAGAGC 2211
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2212 GAGACGAGCTGTGTGAACCGAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTG 2271
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2272 AGCTGGTGTGCGCCCAAGGCGCATCGGCGCAACGAGCAGATCGCAAGCTGTGTAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGlnGlnValAspLysLeuValSer 720
QY 2332 AAGGCGCATCGCAAGTGTCTTCTCGACGCGCATCGAT 2370
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7
US-10-325-468-23
; Sequence 23, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.10 protein Pol
US-10-325-468-23

Alignment Scores:
Pred. No.: 2,38e-161 Length: 1003
Score: 3468.00 Matches: 642
Percent Similarity: 95.28% Conservative: 45
Best Local Similarity: 89.04% Mismatches: 30
Query Match: 75.67% Indels: 4
DB: 16 Gaps: 2

US-09-610-313B-30 (1-2469) x US-10-325-468-23 (1-1003)
QY 220 TTCTTCGCGAGGACCTTGGCTTCCCCAGGCAAGCGCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
QY 280 AACCGCGCCCAACGCGCCCGAGCTGCGAGTG-----CGCGGCGCAACACCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
```



Qy	334	CGCAGCGAGCGCGCGCGGAGCGCGCAGCGCCACCTG-----AACTTCCCCCAGATCACC	387
Db	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr	60
Qy	388	CTGTGGCAGCGCGCCCTGTGTGAGCATCAAGTGTGGCGCGCCAGATCAAGGAGGCGCCCTGCTG	447
Db	61	LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu	80
Qy	448	GACACCGCGCGCAGCAGCACCGTCTCGAGGAGATGAGCCTGCCCGCGCAAGTGAAGCC	507
Db	81	AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyAArgTrpLysPro	100
Qy	508	AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGTGCGCCAGTACGACCCAGATCTGTGATC	567
Db	101	LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProIle	120
Qy	568	GAGATCTCGCGCAGAGAGCCATCGGCACCGTGTGATCGCGCCGCCCGCGTGAACATC	627
Db	121	AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle	140
Qy	628	ATCGCGCGCAACATGCTGACCCAGCTGGGTGACCTGAACTTCCGCATCAGCCCCATC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProLieserProIle	160
Qy	688	GAGACCTGTGCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGTGAAGCAGTGGCCC	747
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
Qy	748	CTGACCGGAGGAGATCAAGGCCCTGACCGCCATCTCGGAGNGATGGAAGGAGGCG	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
Qy	808	AAGATCACCAAGATCGCGCCCGAGAACCCCTCAACACCCCGGTGTGGCCATCAAGAAG	867
Db	201	LysIleSerLysIleGlyProGluAsnProTyraThrValPheAlaIleLysLys	220
Qy	868	AAGCAGCAGCACAAGTGGCGCAAGCTGGTGAAGTTCGCGAGCTGAACAAGCGCACCCAG	927
Db	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln	240
Qy	928	GACTTCTGGGAGTGCAGCTGGGCATCCCCACCCCGCGCTCGAAGAAGAAGAAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
Qy	988	GTGACCGTGTGGAGTGGCGACGCTTACGTGTCGCTTCGACGAGGACTTCCTCCG	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheArg	280
Qy	1048	AAGTACACCGCTTCACATCCCCAGATCAACAACGAGACCCCGCGCATCCGCTACCAG	1107
Db	281	LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyArg	300
Qy	1108	TACAACTGTGTCGCCCGAGGGCTGAAGGCGACGCCAGCATCTTCCAGAGCAGCATCACC	1167
Db	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320
Qy	1168	AAGATCTCGGAGCCCTTCCGCGCCCGCAACCCGAGATCGTGTACTACAGTACATGGAC	1227
Db	321	LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrgIntyrMetAsp	340
Qy	1228	GACCTGTAGTGGGACGGACCTGGAGATCGGCCAGCACCGCGCCAGATCGAGGACCTG	1287
Db	341	AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
Qy	1288	CGCAAGCACCTGTGCTGCGTGGGGCTTCACACCCCGCAAGAAGCACCAGAAGGAGCCC	1347
Db	361	ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
Qy	1348	CCCTTCTGTGGATCGAGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG	1407
Db	381	ProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal	400

QY	1408	CTGCCCGAAGAGGAGAGCTGGACCGTGAACGACATCCAGAGAGCTGGTGGCGCAGCTGAAC	146
DB	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
QY	1468	TGGSCCAGCCAGAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTCTCGCGCGC	1527
DB	421	TrpAlaSerGlnIleTyrAlaGlyIleLysValGlnLeuCysLysLeuLeuArgGly	440
QY	1528	GCCAAAGCCCTGACCGACATCGTGCCCTTGACCGCAGGAGGCCGAGCTGGAGCTGGCCGAG	1587
DB	441	ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu	460
QY	1588	AACCGCAGATCTCTGGCGAGCCCGTGACGGCGTGTACTACCAACCCACGACGACCTG	1647
DB	461	AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu	480
QY	1648	GTGGCCGAGATCCAGAAGCAGGGCCACGACGAGTGAGACTTACCAAGATCTTACCAAGAGCCCC	1707
DB	481	IleValGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro	500
QY	1708	TTCAAGAACTTGAAACCGGCAAGTACGCCAAGATGGCCACGCCCCACCAACGACGCTG	1767
DB	501	PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgSerAlaHisThrAsnAspVal	520
QY	1768	AACGAGCTGACCGAGCCGTCAGAGAGATCGCCATCGAGAGCATCTGTACTCTGGGCAAG	1827
DB	521	LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys	540
QY	1828	ACCCCAAGTTCCGCTGCCCATCCAGAAGAGACCTGGGAGACCTGGTGGACCGACTAC	1887
DB	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr	560
QY	1888	TGGCAGGCCACTGGATCCCGAGTGGGAGTTCTGTGAAACACCCCGCCCTGGTGAAGCTG	1947
DB	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu	580
QY	1948	TGTTACCAGCTGGAGAGAGGCCCATCATCGCGCCGAGACCTTCTACGTGGACGCGGCC	2007
DB	581	TrpTyrGlnLeuGluLysGluProIleValGlyValGluThrPheTyrValAspGlyAla	600
QY	2008	GCCAAACCGCAGACCAAGATCGGCAAGCCGGCTACGTGACCGACCGCGGCCCGCGCAAG	2067
DB	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys	620
QY	2068	ATCGTAGCCTGACCGAGACCAACCAAGAACCGAGCTGCGAGGCCATCCAGCTGGCC	2127
DB	621	ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla	640
QY	2128	CTGCAGCACAGCGCGCAGCGAGTGAACATCGTGACCGCAGCACCGAGTACGCCCTGGGCATC	2187
DB	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle	660
QY	2188	ATTCAGGCCCGCCGCAAGAGCGAGAGCGAGCTGGTGAACACAGATCATCTGAGCAGCTG	2247
DB	661	IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu	680
QY	2248	ATCAAGAAGAGAGGTGTACCTGAGCTGGTGGTCCCGCCACAAAGGCGATCGCGCGCAAC	2307
DB	681	IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn	700
QY	2308	GAGCAGATGCACAAGCTGGTGAGCAAGGGCATCCGCAAGTGTGTCTCTGGACCGCATC	2367
DB	701	GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle	720
QY	2368	GAT 2370	
DB	721	Glu 721	

RESULT 8  
US-10-325-468-35  
; Sequence 35, Application US/10325468  
; Publication No. US2004010823A1  
; GENERAL INFORMATION:

```

; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol
; US-10-325-468-35

Alignment Scores:
Pred. No.: 2,98e-161 Length: 1003
Score: 3466.00 Matches: 643
Percent Similarity: 95.28% Conservative: 44
Best Local Similarity: 89.18% Mismatches: 30
Query Match: 75.63% Indels: 4
DB: 16 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-325-468-35 (1-1003)

Qy 220 TTCTTCGGCAGGACCTGGCTTCCCGCAGGCGGCGAGTTCCTCCCGAGGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerSerGluGln 20
Qy 280 AACCGCGCCCAACGCCACCAGCGCGAGCTGCAGGTG---CGCGCGAC---AACCCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgGlyAspAsnSer 40
Qy 334 CGCAGCGAGGCGCGCGCGAGCGCGAGCGCACCTG-----AACTTCCCGCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
Qy 388 CTGTGGCAGGCGCCCTGTGTGAGATCAAGTGGCGCGCGAGTCAAGAGGCGCTCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
Qy 448 GACACCGCGCGCAGCACCGTCTCGAGGAGATGAGCTGCCCGGCAAGTGAAGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
Qy 508 AAGATGATCGCGGCATCGCGGCTTCATCAAGGTGCGCGAGTACGACGACGATCCTGATC 567
Db 101 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProIle 120
Qy 568 GAGATCTGGCGCAAGAGGCGATCGGACCGTGTGATCGGCGCCCGCCCGCGTGAATC 627
Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
Qy 628 ATCGCGCGCAACATGCTGACCCAGCTGGCTGGACCTGCACTTCCCGCATCGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
Qy 688 GAGACCGTGCCTGTAAGCTGAAGCCCGGATGACCGCGCCCAAGGTGAAGCATGTGCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
Qy 748 CTGACCGAGGAGATCAAGGCGCTGACCGCCATCTCGAGGAGATGGAAGAGGCG 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
Qy 808 AAGATCACCAAGATCGGCGCGGAGAACCCCTACAGACCCCGCTGTCGCGCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

```

```

Qy 868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGAGCTTCCCGAGCTTGAACAACGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240
Qy 928 GACTTCTGGGAGGTGACGTGGCGATCCCGACCCCGCGCGCTGAAGAGAAGAGGAC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
Qy 988 GTGACCGTGTGGACGTGGCGCGACCTACTTACGCTGCCCTGGAGCAGGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
Qy 1048 AAGTACACCGCTTCCACCATCCCGAGCATCAACAACGAGACCCCGCGCATCCGCTAC 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
Qy 1108 TACAAGCTGTGCTCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
Qy 1168 AAGATCCTGGAGCGCTTCCCGCGCCGCAACCCCGAGATCGTGATCTACCATGATGAC 1227
Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340
Qy 1228 GACTGTGTGCGGACGACCTGGAGATCGGCGAGCATCCCGCGCAAGATCGAGAGCTG 1287
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
Qy 1288 GCGAAGCACCTGTGCTGGCGCTTCCACCGCCCGAGAGACGAGACGAGAGGAGCC 1347
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
Qy 1348 CCCTCTCTGTGGATGGGTACGAGCTCACCGCACAAAGTGGACCGTGCAGCCCATCGAG 1407
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
Qy 1408 CTCCCGGAGAGGAGTGGACCGTGAACACGACATCCAGAACGTGTGGGCAAGCTGAAC 1467
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
Qy 1468 TGGCGCGAGCATCTACCCCGCGCATCAAGTGGCGCGAGTGTGCAAGCTGTGCAAGCTG 1527
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
Qy 1528 GCCAAGCCCTGACCGACATCGTCCCTGACCGAGGAGGCGAGCTGGAGCTGGCGGAG 1587
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuLysLysGlu 460
Qy 1588 AACCGGAGATCTGCGCGAGCGCGTGCACGCGGTGTACTACGACCCCGAGAGGACTG 1647
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
Qy 1648 GTGGCGGAGATCCAGAGCAGGCGCCACGACGAGTGGACCTACCGAGTCTACCGAGCC 1707
Db 481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500
Qy 1708 TTCAAGAACCTGAAGACCGCAAGTACGCCAAGATGGCCACCGCCACACCAACGAGCTG 1767
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520
Qy 1768 AAGCAGCTGACCGAGGCGGTGCAGAAAGATCCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
Qy 1828 ACCCGCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
Qy 1888 TGCAGCGCACCTGGATCCCGGAGTGGAGTTCGTGAACACCCCGCTGGTGGTGAAGCTG 1947
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

```

```
QY 1948 TGGTACCAGCTGGAGAGAGCCCATCATCGCGCGCGAGCTTCTACGTGGAGCGCGCC 2007
Db 581 TrpTyrGlnLeuGluLysGluProlleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2008 GCCAACCGCGAGACCAAGATCGCAAGCGCGGTACGTGACCCAGCCGGGCGCGAGAAG 2067
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620
QY 2068 ATCGTAGCCTGACCGAGACCAACACAGACCGAGTGCAGGCATCCAGCTGGCC 2127
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaHisLeuAla 640
QY 2128 CTCAGAGACCGCGAGCGAGGTGAACATCGTGACCGACAGCGAGTACGCGCTGGGCATC 2187
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGGCCCGCCGACCAAGACGAGCGAGCTGGTGAACACAGATCATCGAGCGCTG 2247
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCGCCGCCACAGGGCATCGCGGCAC 2307
Db 681 IleLysLysGluLysValTyrLeuThrTrpIleProAlaHisLysGlyIleGlyGlyAsn 700
QY 2308 GAGCAGATCGACAGCTGGTGAGCAAGGGCATCGCAAGGTGCTGTCTGGAGCGGCATC 2367
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db 721 Glu 721

RESULT 9
US-10-325-468-9
; Sequence 9, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Shibata, Riri
; APPLICANT: Pekrun, Katja
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;
; OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;
; OTHER INFORMATION: clone P10.26 protein Pol
US-10-325-468-9

Alignment Scores:
Pred. No.: 3,33e-161 Length: 1003
Score: 3465.00 Matches: 642
Percent Similarity: 95.15% Conservative: 44
Best Local Similarity: 89.04% Mismatches: 31
Query Match: 75.61% Indels: 4
DB: 16 Gaps: 2

US-09-610-313B-30 (1-2469) x US-10-325-468-9 (1-1003)
QY 220 TTCTTCGCGAGGACCTGGCGCTTCCCGCAGGGCAGCGCGCGAGTTCGCCAGGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
```

```
QY 280 AACCGCGCAACAGCGCCCGCCACCGAGCGCGAGCTGCAGGTG-----CGCGCGCAACAACCCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTipArgArgAspAsnAsnSer 40
QY 334 CGCAGCAGGCGCGCGCGCCAGCGCCAGCGCACCGCTG-----AACTTCCCGCAGATCACCC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCGCCCTGTGTGAGCATCAAGTGGGGCGGCAGATCAAGAGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGGGCGCCGACACACCGCTGTGTGAGAGAGATGAGCTGCCCGCAAGTGAAGACCCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTyrLysPro 100
QY 508 AGATGATCGGGCGCATCGCGCGCTTTCATCAAGGTGGCGCAGTACACCGAGTACCTGATC 567
Db 101 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120
QY 568 GAGATCTGGCGCAAGAGGCCATCGGCACCGCTGTGATCGGCCCGCCACCGCTGAACATC 627
Db 121 AspileCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCATCAGCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGAAGCTGAAGCGCGCGCATGAGCGCGCGCCCGCCAAAGTGAAGCAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGGCGAGAGATGGAAGAGGAGGCG 807
Db 181 LeuThrGluGlnLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGCGAGAACCCCTACAACACCCCGCTGTTCGCCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGACCAAGTGGCGCAAGCTGTGTGGTTCGCGGAGCTGAACAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240
QY 928 GACTTCTGGGAGGTGCAGCTGGCGCATCCCCACCGCGCGCTGAAGAGAGAGAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGAGACGTGGCGCGACGCTACTTTCAGCGTGCCTGAGCAGAGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCTTCCATCCCGCAGCATCAACAGAGACCCCGCGCATCCGCTACCGAC 1107
Db 281 LysThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAAGTGTGCGCCAGCGCTCGAAGGCGAGCCCGCAGCATCTTCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCCTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCGTGTATCAAGTACATGGAC 1227
Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340
QY 1228 GACCTGTAGTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCCAGATCGAGGAGCTG 1287
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1288 CGCAAGCACCTGTGTGGCTGGGCTTCACCCACCGCGCACCAAGAGAGAGAGAGAGAGCC 1347
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1348 CCCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAG 1407
```

Db	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal	400
QY	1408	CTGCCCCGAGAGAGAGCTGGACCGTGAACACACATCCAGAAAGCTGGTGGGCGACAGCTCAAC	1467
Db	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
QY	1468	TGGGCCACCCAGATCTACCCCGGCATCAAGGTGCGCAGCTGTGCACAGCTGCTCGCGCGC	1527
Db	421	TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly	440
QY	1528	GCCAAGGCCCTGACCGACATCGTCCCTTGACCCAGGAGGCCGAGCTGGAGCTGGCCGAG	1587
Db	441	ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu	460
QY	1588	AACCGCCAGATCTCTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCACGACCAAGGACCTG	1647
Db	461	AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu	480
QY	1648	GTGCCCGAGATCCAGAAGACGGGCCACGACCCAGTGGACCTACCAAGATCTACCGAGAGCCC	1707
Db	481	IleValGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro	500
QY	1708	TTCAAGAACCTGAAGACGGCAAGTAGCCCAAGATGCGCACCGGCCACCAACGACGCTG	1767
Db	501	PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal	520
QY	1768	AAGCAGCTGACCGAGGCGTGCAGAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG	1827
Db	521	LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys	540
QY	1828	ACCCCCAAGTTCCGCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACTAC	1887
Db	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr	560
QY	1888	TGGCAGGCCACCTGGATCCCGAGTGGAGTTCTGTGAACACCCCCCCCCCTGGTGAGCTG	1947
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu	580
QY	1948	TGTTACCAAGCTGGAGAAGGACCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC	2007
Db	581	TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla	600
QY	2008	GCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGCGCAGAAG	2067
Db	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys	620
QY	2068	ATCGTGAGCTGACCGAGACCAACCAACAGAAGACCGAGCTGCGAGGCCATCCACAGCTGCC	2127
Db	621	ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla	640
QY	2128	CTGCAGACAGCGCGACGAGGTGAACATCGTGAACCGACGACCGACGTACGCCCTGGGCATC	2187
Db	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle	660
QY	2188	ATCCAGGCCACCGCCGACCAAGACGACGAGCTGGTGAACACAGATCATCGACGACGCTG	2247
Db	661	IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu	680
QY	2248	ATCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCCCGCCCAACAGGCGCATCGCGCGCAAC	2307
Db	681	IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn	700
QY	2308	GAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGCAAGGTGTGTTCTCTGGAGAGGCATC	2367
Db	701	GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle	720
QY	2368	GAT 2370	
Db	721	Glu 721	
RESULT 10			
US-10-325-468-20			

```

; Sequence 20, Application US/10325468
; Publication No. US20040101822A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Kiri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20

Alignment Scores:
Pred. No.: 3.33e-161 Length: 1003
Score: 3465.00 Matches: 642
Percent Similarity: 95.15% Conservative: 44
Best Local Similarity: 89.04% Mismatches: 31
Query Match: 75.61% Indels: 4
DB: 16 Gaps: 2

```









Db 594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613  
Qy 2071 GTGACCTGACCGAGACACCAACAGACCGAGCTGCAGGCCATCCAGCTGCCCCGTG 2130  
:::|||||  
Db 614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633  
Qy 2131 CAGACAGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGCATCATC 2190  
:::|||||  
Db 634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653  
Qy 2191 CAGCCAGCCGACAGAGCGAGCGAGCTGTGTGAACACCATCATCGACGCGTCATC 2250  
:::|||||  
Db 654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673  
Qy 2251 AAGAAGAGAGAGGTGTACCTGAGCTGGGTGCCGCCGCCACAGGGCATCGCGGCAACGAG 2310  
:::|||||  
Db 674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693  
Qy 2311 CAGATCGACAAGCTGGTGCAGCAAGGGCATCGCGAAGGTGCTGCTCGGACGGCATCGAT 2370  
:::|||||  
Db 694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713

RESULT 13  
US-10-296-734-1471  
; Sequence 1471, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A  
; TITLE OF INVENTION: Synthetic molecules and uses therefor  
; FILE REFERENCE: Savine  
; CURRENT APPLICATION NUMBER: US/10/296,734  
; CURRENT FILING DATE: 2003-08-04  
; PRIOR APPLICATION NUMBER: AU PQ7761/00  
; PRIOR FILING DATE: 2000-05-26  
; NUMBER OF SEQ ID NOS: 1507  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1471  
; LENGTH: 1006  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: POL Consensus A  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (15)..(15)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (22)..(23)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (26)..(26)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (30)..(30)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (36)..(36)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (38)..(39)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (41)..(44)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (331)..(333)

; LOCATION: (46)..(46)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (48)..(48)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (51)..(51)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (53)..(55)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (58)..(58)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (73)..(73)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (75)..(75)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (105)..(105)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (130)..(130)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (173)..(173)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (194)..(194)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (197)..(197)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (255)..(255)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (280)..(281)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (293)..(293)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (300)..(300)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (316)..(316)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (320)..(320)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (331)..(333)

```
;
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (335)..(335)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (365)..(365)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (369)..(369)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (373)..(373)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (402)..(403)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (434)..(434)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (439)..(439)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (452)..(452)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (470)..(470)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (480)..(480)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (487)..(487)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (514)..(514)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (534)..(535)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (544)..(544)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (553)..(553)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (593)..(593)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (610)..(610)
; OTHER INFORMATION: X = unknown amino acid
```

```
;
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (675)..(675)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:

Alignment Scores:
Pred. No.: 1,23e-152 Length: 1006
Score: 3289.50 Matches: 622
Percent Similarity: 89.36% Conservative: 25
Best local Similarity: 85.91% Mismatches: 70
Query Match: 71.78% Indels: 7
DB: 15 Gaps: 2

US-09-610-313b-30 (11-2469) x US-10-296-734-1471 (11-1006)

QY 220 TTCTTCGGGAGGAGCTGGCTTCCCGGAGGGAAGCCCGGAGTTCCTCCCGAGGAGCAG 279
Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArg***PheSerSerGluGln 20
QY 280 AACCGCGCCAAACAGACCCACCGCGGAGCTGCAGTGCAGCGCGACACACCC--- 333
Db 21 Thr*****AsnSer***ThrSerArg***LeuTrpAspGlyGly***Asp*****Leu 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCCAGGGC-----ACCTGAACTTCCCC 378
Db 41 *****Gly***Glu***GlnGly***Gly*****SerPhe***PhePro 60
QY 379 CAGATCACCTGTGGCAGGCGCCCGCTGGTGAGCATCAAGTGGCGGCGGCGGATCAAGGAG 438
Db 61 GlnIleThrLeuTrpGlnArgProLeuValThrVal***Ile***GlyGlnLeuIleGlu 80
QY 439 GCCTCTGTGACACCGCGCGCGGAGGACACCGCTGCTGGAGGAGATGAGCTGCCCGGCAAG 498
Db 81 AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLys 100
QY 499 TGAAGCCCAAGATGATCGCGCGCATCGCGGGTTTCATCAAGGTGCGCCAGTAGCAGACCAG 558
Db 101 TrpLysProLys***IleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGln 120
QY 559 ATCTGTATCGAGATCTCGCGCAGAGAGCCATCGGCACCGTGTGATCGGCCCGCCACCCC 618
Db 121 IleLeuIleGluIleCysGlyLysLys***IleGlyThrValLeuValGlyProThrPro 140
QY 619 GTCAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGACTTCCCCATC 678
Db 141 ValAsnIleIleGlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160
QY 679 AGCCCATCGAGACCGTGCCTGGAAGCGGATGAAAGCCCGGATGAGCGGCCCAAGGTGAAG 738
Db 161 SerProIleGluThrValProValLysLeuLysPro***MetAspGlyProLysValLys 180
QY 739 CAGTGGCCCTGACCGGAGGAGATCAAGGCCCTGACCGCCATCTCGGAGGAGATGAG 798
Db 181 GlnTrpProLeuThrGluGluLysIleLysAlaLeuThr***IleCys***GluMetGlu 200
QY 799 AAGGAGGCGCAAGATCACCAAGATCGCGCGCGAGAACCCCTACACACCCCGTGTTCGCC 858
Db 201 LysGluGlyLysIleSerLysIleGlyProGluAsnProTyrrAsnThrProValPheAla 220
QY 859 ATCAAGAAGAGGACAGCACCAAGTGGCGGCAAGCTGGTGAGCTTCGCGGAGCTGAACAAG 918
Db 221 IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 240
QY 919 CGCACCCAGGACTTCTGGGAGGTGCAAGTGGGATCCCCACCCCGCCGCTGAAGAAG 978
Db 241 ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis***AlaGlyLeuLysLys 260
QY 979 AAGAAGAGCGTGACCGTGTGGACGTGGGCGAGCCCTACTTTCAGCGTCCCGCTGGAGCAG 1038
Db 261 LysLysSerValThrValLeuAspValGlyAspAlaTyrrPheSerValProLeuAsp*** 280
QY 1039 GACTTCGCAAGTACACCGCTTTCACCATCCCGGATCAACACGAGACCCCGGCGCATC 1098
```

```
Db 281 ***PheArgLysTyrThrAlaPheThrIleProSer***AsnAsnGluThrProGly*** 300
Qy 1099 CGCTACCAAGTACAACGCTGCTCCCGCAGGCTGGAGGCGACCCCGACGATCTTCCAGAGC 1158
Db 301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro***IlePheGln*** 320
Qy 1159 AGCATGACCAAGATCTCTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCTACCAAG 1218
Db 321 SerMetThrLysIleLeuGluProPheArg*****Pro***IleValIleTyrGln 340
Qy 1219 TACATGACGACCTGTACGTGGGCGAGCGACCTGGAGATCGCGCCAGCCGCGCGCAAGATC 1278
Db 341 TyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360
Qy 1279 GAGGAGCTGCCAAGCACCTCTCGCTGGTGGGCTTCCACACCCCGCAAGAACGACCCAG 1338
Db 361 GluGluLeuArg***HisLeuLeu***TrpGlyPhe***ThrProAspLysLysHisGln 380
Qy 1339 AAGGAGCCCTCTCTGTGGATGGGTACGAGCTGCACCCCGCAAGAGTGACCGTGCAG 1398
Db 381 LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGln 400
Qy 1399 CCGATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGC 1458
Db 401 Pro*****LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420
Qy 1459 AAGCTGAACCTGGGCGAGCATCTACCCCGGCTCAAGGTGGCGCCAGCTGTCAAGCTG 1518
Db 421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys***LysGlnLeuCys***Leu 440
Qy 1519 CTGCGCGCGCCAAAGGCGCTGACGACATCGTGCCTGCTACCGAGGAGCGCGAGCTGAG 1578
Db 441 LeuArgGlyAlaLysAlaLeuThrAspIleVal***LeuThrGluGluAlaGluLeuGlu 460
Qy 1579 CTGCGCGAGAACCGGAGATCTCTGGGAGGCGCTGCACGCGCTGTACTAGACCCGAGC 1638
Db 461 LeuAlaGluAsnArgGluLeuLys***ProValHisGlyValTyrTyrAspPro*** 480
Qy 1639 AAGCACTGTGGCGGAGATCCAGAAGCAGGCGCCACGACCATGAGCTGACCATCTAC 1698
Db 481 LysAspLeuValAlaGlu***GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500
Qy 1699 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCACCGCCACAC 1758
Db 501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla***LysArgSerAlaHisThr 520
Qy 1759 AACGACGTGAAGAGCTGACCGGCGCTGCAGAAATCGCCATCGGAGCATCGTGATC 1818
Db 521 AsnAspValLysGlnLeuThrGluValValGlnLysVal*****GluSerIleValIle 540
Qy 1819 TGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1878
Db 541 TrpGlyLys***ProLysPheArgLeuProIleGln***GluThrTrpGlu***TrpTrp 560
Qy 1879 ACCGACTACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTG 1938
Db 561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580
Qy 1939 GTGAAGCTGTGTACGAGTGGAGAGGAGCCCATCATCGCGCGGAGACCTTCTACGTG 1998
Db 581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle***GlyAlaGluThrPheTyrVal 600
Qy 1999 GACGGCGCGCCAAACCGGAGACCAAGATCGCAAGCGCGGCTACGTGACCGCGCGGGC 2058
Db 601 AspGlyAlaAlaAsnArgGluThrLys***GlyLysAlaGlyTyrValThrAspArgGly 620
Qy 2059 CGGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTGCAGCGCCATC 2118
Db 621 ArgGlnLysValValSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIle 640
Qy 2119 CAGCTGGCCCTGCAGGACGCGGAGGTGAACATCTGTGACCGGACGACCATGACGCC 2178
```

```
Db 641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660
Qy 2179 CTGGGCGATCATCCAGGCGCCAGCCGACAAAGAGCGAGAGCTGGTGAACCAAGATCATC 2238
Db 661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu***ValAsnGlnIleIle 680
Qy 2239 GAGCAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGGTGCGCGCCCAAGGGCATC 2298
Db 681 GluLysLeuIle***Lys***LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700
Qy 2299 GGCGGCAACGAGCAGATCGCAACAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTGTCTCTG 2358
Db 701 GlyGlyAsnGluGlnValAspLysLeuValSer***GlyIleArgLysValLeuPheLeu 720
Qy 2359 GACGCGCATCGAT 2370
Db 721 AspGlyIleAsp 724

RESULT 14
US-10-301-661A-6
; Sequence 6, Application US/10301661A
; Publication No. US20030157660A1
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-
; APPLICANT: INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAULIERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSI, Francois
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/10/301,661A
; PRIOR FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US/09/319,588C
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRP
; ORGANISM: Human immunodeficiency virus type 1
US-10-301-661A-6

Alignment Scores:
Pred. No.: 2,56e-149 Length: 1014
Score: 3221.50 Matches: 587
Percent Similarity: 90.80% Conservative: 74
Best Local Similarity: 80.63% Mismatches: 56
Query Match: 70.29% Indels: 11
DB: 14 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-301-661A-6 (1-1014)
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGCAAGGCCCGAGATTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgGluLeuTrpValSerGlyGluGlu 20
Qy 280 AAC-----CGCGCAACAGCCCAACCCAGCGAGCGCGCGAGCTGCAGGTTG----- 318
Db 21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGluGlu 40
Qy 319 ---CGCGCGCAACACCCCGCAGCGAGCGCGCGCGAGCGCCAG-----GGCACC 366
Db 41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60
Qy 367 CTGAACTTCCCGAGATCACCCTGTGGCAGCGCCCTGTGTGATCATCAAGTGGCGGC 426
Db 61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
```



; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
 ; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  
 ; FILE REFERENCE: 20747Y  
 ; CURRENT APPLICATION NUMBER: US/09/952,060  
 ; PRIOR FILING DATE: 2001-09-14  
 ; PRIOR APPLICATION NUMBER: PCT/US01/28861  
 ; PRIOR FILING DATE: 2001-09-14  
 ; PRIOR APPLICATION NUMBER: 60/317,814  
 ; PRIOR FILING DATE: 2001-09-07  
 ; PRIOR APPLICATION NUMBER: 60/279,056  
 ; PRIOR FILING DATE: 2001-03-27  
 ; PRIOR APPLICATION NUMBER: 60/233,180  
 ; PRIOR FILING DATE: 2000-09-15  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 35  
 ; LENGTH: 1350  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Codon optimized gag-1A pol. fusion  
 US-09-952-060-35

Alignment Scores:  
 Pred. No.: 4,11e-145 Length: 1350  
 Score: 3135.50 Matches: 600  
 Percent Similarity: 81.29% Conservative: 43  
 Best Local Similarity: 75.85% Mismatches: 60  
 Query Match: 68.42% Indels: 88  
 DB: 10 Gaps: 5

US-09-610-313B-30 (1-2469) x US-09-952-060-35 (1-1350)

Qy 13 ATGCGCCGAGCCATGAGCCAGGCCACC---AGCGCCACATCTCTGATGCGAGCGCAGCAAC 69  
 Db 363 LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArggIysn 382  
 Qy 70 TTCAAGGGCCCAAGCCATCATCAAGTGTCTTCAACTGCGCGCAAGAGGAGGCCACATCGCC 129  
 Db 383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysValGlyHisIleAla 402  
 Qy 130 CGCAACTGCGCGCCCGCCCGCAAGAAGGGTGTCTGGAAGTGTGCGCAAGAGAGGGCCACAG 189  
 Db 403 LysAsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422  
 Qy 190 ATGAAGGACTGCACCGAGCGCCAGCCCACTTCTCCGCGAGGACCTGGCTTCCCGCCAG 249  
 Db 423 MetLysAspCysAsnGluAa-gglnAlaAsnPheLeu-GlyLysIleTrpProSerHisLy 442  
 Qy 250 GGCAGGCGCGAGTTCCTCCAGCGAGCAGAACCGCGCCCAACAGCCCGCCAGCGCGAG 309  
 Db 442 sGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProGluGluSe 462  
 Qy 310 CTGAGTGTGGCG-----GGGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGG 363  
 Db 462 rPheArgPheGlyGluGluLysThrThrProSerGlnLysGlnGluProIleAspLys-- 481  
 Qy 364 ACCTGTAACCTCCCGCAGATCACCTGTGGCAGCGCCCTCTGGTGTGACATCAAGTGGGC 423  
 Db 482 GluLeu-----TyrProLeuAlaSerLeuAa----- 490  
 Qy 424 GGCCAGATCAAGAGGGCCCTGTGGACACACCGCGCCGACGACACCGTGTGGAGGATG 483  
 Db 490 ----- 490  
 Qy 484 AGCCTGCCCGCAAGTGGAAAGCCCAAGATGATCGCGGCATCGCGCGCTTTCATCAAGGTG 543  
 Db 490 ----- 490  
 Qy 544 CGCCAGTACGACCCAGATCCTGATCGAGATCTGGGGCAAGAGGCCATCGGCACCGTCTG 603  
 Db 490 ----- 490

Qy 604 ATCGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGTCACC 663  
 Db 491 -----SerLeuPheGlyAsnAspProSerSerGlnMetAla----- 502  
 Qy 664 CTGNACTTCCCATCAGCCCATCAGACCGTGCCTGCGTGAAGCTGAAGCCCGGATGAC 723  
 Db 503 -----ProfileSerProIleGluThrValProValLysLeuLysProGlyMetAsp 519  
 Qy 724 GGCCTCAAGTGAAGAGTGGCCCTGACCGAGAGAGATCAAGGGCCCTGACCGCCATC 783  
 Db 520 GlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIle 539  
 Qy 784 TGCAGAGATGGAAGAGAGGGCAAGATCACCAAGATCGCCCGGAGAACCCCTACAAAC 843  
 Db 540 CysThrGluMetGluLysGluLysIleSerLysIleGlyProGluAsnProTyrAsn 559  
 Qy 844 ACCCGCTGTCGCCATCAAG 903  
 Db 560 ThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPhe 579  
 Qy 904 CGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGGCATCCCCACCCC 963  
 Db 580 ArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPro 599  
 Qy 964 GCGGCTGGAAG 1023  
 Db 600 AlaGlyLeuLysLysLysSerValThrValLeuAlaValGlyAspAlaTyrPheSer 619  
 Qy 1024 GTGCCCTGACGAGGACTTCCGCAAGTACACCGCTTACCCTCCAGCATCAACAAAC 1083  
 Db 620 ValProLeuAspGluAspPheArgLysThrAlaPheThrIleProSerIleAsnAsn 639  
 Qy 1084 GAGACCCCGGCTCCGCTACAGTACAACTGCTGCCCGCCAGGGCTGGAAGGGCAGCCCC 1143  
 Db 640 GluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro 659  
 Qy 1144 AGCATCTTCAG 1203  
 Db 660 AlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAsp 679  
 Qy 1204 ATCGTATCTACAGTATGAG 1263  
 Db 680 IleValIleTyrGlnTyrMetAlaLeuLysValGlySerAspLeuGluIleGlyGln 699  
 Qy 1264 CACGCGCCCAAGATCGAGAGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1323  
 Db 700 HisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrPro 719  
 Qy 1324 GACAAG 1383  
 Db 720 AspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAsp 739  
 Qy 1384 AGTGAAG 1443  
 Db 740 LysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIle 759  
 Qy 1444 CAGAAGTGTGTGGCAAGTGAAGTGGCCAGCAGATCTACCCCGCATCAAGAGTGGC 1503  
 Db 760 GlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArg 779  
 Qy 1504 CAGTGTGCAAGTGTGTGCGCGGCCCAAGGCCCTGACCGATCGTGCCTCCCTGACCGAG 1563  
 Db 780 GlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGlu 799  
 Qy 1564 GAGCGAGTGTGAGTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1623  
 Db 800 GluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVal 819  
 Qy 1624 TACTACGACCCAG 1683  
 Db 820 TyrTrpAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrp 839



QY 1684 ACCTACCAGATCTACACGAGCCCTTCAAGAACTGAAGACCGCAAGTACGCAAGATG 1743  
 Db |||||||TyrGlnIleTyrGlnIlePheLeuLysThrGlyLysThrAlaArgmec 859  
 QY 1744 CGCACCGCCACACCAACGAGCTGACCGAGCCCTGCGAAGATCGCCATG 1803  
 Db |||||||ThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr 879  
 QY 1804 GAGAGCATCTGATCTGGGCAAGACCCCAAGTTCGGCTGCCATCCAGAGGAGACC 1863  
 Db |||||||ValIleTyrGlyLysThrProLysPheLysLeuProLysGlnLysGluThr 899  
 QY 1864 TGGGAGACCTGGTGGACCCGACTACTGGCAGCCACCTGGATCCCGAGTGGGAGTTCGTG 1923  
 Db |||||||TrpThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVal 919  
 QY 1924 AACACCCCCCTTGTGAGCTGTGTACCGCTGGAGAGGAGCCCATCATCGGCGCC 1983  
 Db |||||||AsnThrProProLeuValLysLeuTyrTrpGlnLeuGluLysGluProIleValGlyAla 939  
 QY 1984 GAGACCTTCTACGTGACGCGCGCCCAACCGGAGACCAAGATCGCGAGCGCGCTAC 2043  
 Db |||||||GluThrPheTyrValAlaGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyr 959  
 QY 2044 GTGACGACCGGCGCGGAGAGATCGTGCAGCTGACCGAGACCAACCAAGAGACC 2103  
 Db |||||||ValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThr 979  
 QY 2104 GAGCTGAGGCCCATCCAGCTGGCCCTGCGAGACGCGGCGAGGAGTGAAACATCGTGACC 2163  
 Db |||||||AlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThr 999  
 QY 2164 GACAGCCAGTACGCCCTGGGCATCATCCAGCGCCGAGCCCGACAGAGCGAGCGAGCTG 2223  
 Db |||||||AlaSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeu 1019  
 QY 2224 GTGAACCCAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGTGCCC 2283  
 Db |||||||ValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValPro 1039  
 QY 2284 GCCCACAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGC 2343  
 Db |||||||AlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArg 1059  
 QY 2344 AAGTGCTGTCTCTGGACGGCATCGAT 2370  
 Db |||||||LysValLeuPheLeuAspGlyIleAsp 1068

Search completed: June 2, 2005, 05:27:02  
 Job time : 245.54 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 20:20:55 ; Search time 209.175 Seconds  
(without alignments)  
12088.668 Million cell updates/sec

Title: US-09-610-313B-30  
Perfect score: 4583  
Sequence: 1 gtcagccaccatggcga.....gggctagaccggtgaattc 2469

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2.1/USPTO.epool/US09610313/runat\_31052005.155136.15147/app.query.fasta.1.7893  
-DB-UniProt\_03 -QWMT-fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US09610313 @CGN 1 1 725 @runat\_31052005.155136.15147 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4178	91.2	1427	2 Q9WF92	Q9wf92 human immun
2	4174	91.1	1427	2 Q9WF89	Q9wf89 human immun
3	4170	91.0	1428	2 Q9WF86	Q9wf86 human immun
4	4028	87.9	1427	2 Q9WF62	Q9wf62 human immun
5	4021	87.7	1429	2 Q9X4F8	Q9x4f8 human immun
6	3993	87.1	1425	2 Q9X4Q6	Q9x4q6 human immun
7	3986	87.0	1427	2 Q9WF96	Q9wf96 human immun
8	3982	86.9	1420	2 Q9X4R4	Q9x4r4 human immun
9	3973.5	86.7	1427	2 Q9WF71	Q9wf71 human immun
10	3969	86.6	1427	2 Q9WF80	Q9wf80 human immun
11	3966.5	86.5	1427	2 Q9WF65	Q9wf65 human immun
12	3965	86.5	1437	2 Q9WF77	Q9wf77 human immun
13	3957.5	86.4	1427	2 Q9WF74	Q9wf74 human immun
14	3951.5	86.2	1427	2 Q9ADX5	Q9adx5 human immun
15	3943	86.0	1425	2 Q9WF79	Q9wf79 human immun
16	3936	85.9	1427	2 Q9WF98	Q9wf98 human immun

#### ALIGNMENTS

##### RESULT 1

17	3931.5	85.8	1429	2	Q9WF84	Q9wf84 human immun
18	3922	85.6	1428	2	Q6X4P0	Q6x4p0 human immun
19	3912.5	85.4	1427	2	Q9WF68	Q9wf68 human immun
20	3876.5	84.6	1432	2	Q8ADZ9	Q8adz9 human immun
21	3873	84.5	1433	2	Q8AD89	Q8ad89 human immun
22	3871	84.5	1434	2	Q8ADV1	Q8adv1 human immun
23	3869	84.4	1430	2	Q7SV36	Q7sv36 human immun
24	3865	84.3	1425	2	Q7SV20	Q7sv20 human immun
25	3859.5	84.2	1433	2	Q8ADN7	Q8adn7 human immun
26	3858.5	84.2	1433	2	Q8ADK5	Q8adk5 human immun
27	3856.5	84.1	1433	2	Q8ADS7	Q8ads7 human immun
28	3853.5	84.1	1433	2	Q8ADG5	Q8adg5 human immun
29	3850.5	84.0	1435	2	Q9DQ33	Q9dq33 human immun
30	3849.5	84.0	1433	2	Q8ADL1	Q8adel1 human immun
31	3844.5	83.9	1431	2	Q8ADR1	Q8adr1 human immun
32	3843.5	83.9	1433	2	Q8ADCS	Q8adcs human immun
33	3842.5	83.8	1433	2	Q8AE32	Q8ae32 human immun
34	3841.5	83.8	1433	2	Q8ADM1	Q8adml1 human immun
35	3841.5	83.8	1433	2	Q8ADU3	Q8adu3 human immun
36	3840.5	83.8	1437	2	Q41782	Q41782 human immun
37	3839	83.8	1434	2	Q8QW727	Q8qw727 human immun
38	3839	83.8	1435	2	Q9WLJ3	Q9wlj3 human immun
39	3838.5	83.8	1433	2	Q8AD85	Q8ad85 human immun
40	3836.5	83.7	1433	2	Q7SPT9	Q7spt9 human immun
41	3832.5	83.6	1433	2	Q8ADW7	Q8adw7 human immun
42	3832.5	83.6	1436	2	Q8AD93	Q8ae23 human immun
43	3830.5	83.6	1433	2	Q8AE23	Q8ae23 human immun
44	3829.5	83.6	1433	2	Q8ADR9	Q8adr9 human immun
45	3827.5	83.5	1433	2	Q8ADH3	Q8adh3 human immun

Q9WF92 PRELIMINARY; PRT; 1427 AA.  
AC Q9WF92;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OC NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana";  
RT J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110975; AAD17135.1; -;  
DR HSSP; P04585; 1JLA.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0010574; P:DNA integration; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.



Db	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu	879
Qy	1573	CTGAGCTGGCGAGAACCGCGAGATCCTGCGGAGCCCGTGCACGCGGTGTACTACGAC	1632
Db	880	LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp	899
Qy	1633	CCGAGCAAGNACTGGTGGCCGAGATCCAGNAGCAGGCCACGACGAGCTGGACTTACCAG	1692
Db	900	ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrrpThrTyrGln	919
Qy	1693	ATCTTACCAGAGCCCTTCAAGAACTCAAGAACCGGCAAGTAGTACGCCAAGATGCGCACCGCC	1752
Db	920	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
Qy	1753	CACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCGCCATGAGAGCATC	1812
Db	940	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle	959
Qy	1813	GTGATCTGGGGCAAGACCCCCAAGTTCGCGTCCCATCCAGAAGGAGACCTCGGAGACC	1872
Db	960	ValIleTrrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrrpGluThr	979
Qy	1873	TGTTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTTCGTGAACACCCCC	1932
Db	980	TrrpTrrpThrAspTyrTrrpGlnAlaThrTrrpileProGluTrrpGluPheValAsnThrPro	999
Qy	1933	CCCTGTGTGAAGCTGTGTGTACCACTGGAGAGGAGGCCATCATCGCGCGGAGACCTTC	1992
Db	1000	ProLeuValLysLeuTrrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe	1019
Qy	1993	TACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGAC	2052
Db	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp	1039
Qy	2053	CGGGCCCGCAGAAAGATCGTAGCCTGACCCGAGACCAACCAAGAGACCGAGCTGCAG	2112
Db	1040	ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln	1059
Qy	2113	GCCATCCAGCTGGCCCTGCAGGACAGCGCGCAGCGAGTGAACATCGTGACCGCAGACCCAG	2172
Db	1060	AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln	1079
Qy	2173	TACGCCCTGGGCATCATCCAGGCCCCCGCACAGACGAGCGAGCTGGTGAACACAG	2232
Db	1080	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln	1099
Qy	2233	ATCATCGAGCAGCTGATCAAGAGGAGAAAGGTGTACCTGAGCTGGTGGCCGCCACAG	2292
Db	1100	IleIleGlnGlnLeuIleLysLysGluLysValTyrLeuSerTrrpValProAlaHisLys	1119
Qy	2293	GGCATCGCGCGCAACGACGACATCGACAAGCTGGTGAGCAAGGCGCATCCGCCAAGGTGCTG	2352
Db	1120	GlyIleGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu	1139
Qy	2353	TTCTTGACCGGCATCGAT	2370
Db	1140	PheLeuAspGlyIleAsp	1145

KW ARDS; Asparaginyl protease; Core protein; Hydrolase; Metal-binding;  
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
KW Zinc-finger.  
SQ SEQUENCE 1427 AA; 161293 MW; 270F25DF67615C06 CRC64;

Db 360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379  
:::|||||  
Qy 73 AAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGCGCAAGGAGGCGCCACATCGCCGC 132  
Db 380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
Qy 133 AACTGCGCGCCCCCCCAAGAGGGCTGCTGGAAGTGGCGCAAGGAGGCGCCACCATG 192  
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
Qy 193 AAGACTGCACCGAGCGCCAGGCCAACTTCTTCGGGAGGACCTGGCTTCCCCAGGCG 252  
Db 420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439  
Qy 253 AAGCGCGCGAGTCTCCCGAGGACAGAACCGCGCCACAGCCCCCACCAGCGCGAGCTG 312  
Db 440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459  
Qy 313 CAGGTGGCGCGGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 372  
Db 460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479  
Qy 373 TTCCCCCAGATCACCTGTGGCAGCGCCCTCTGTGAGCATCAAGGTGGCGCGCCAGATC 432  
Db 480 PheProGlnIleThrLeuTyrGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499  
Qy 433 AAGGAGCCCTGTGGACACCGCGCGCGAGCACACCGTGTGGAGGAGATGAGCCTGCC 492  
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519  
Qy 493 GGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCAAGGTGGCGCGAGTAC 552  
Db 520 GlyLysTrpLysProLysMetIleGlyIleGlyPheIleLysValArgGlnTyr 539  
Qy 553 GACCAGATCTCGAGATCTCGGCAAGAAGGCCATCGGCACCGTGTGATCGGCCCC 612  
Db 540 AspGlnIleLeuIleGluIleCysGlyLysAlaIleGlyThrValLeuIleGlyPro 559  
Qy 613 ACCCGGTGAACATCATCGCGCGCGCACTGCTACCCAGCTGAGGAGGCGCGCGCGCGCG 672  
Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579  
Qy 673 CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCGCGCATGAGCGCGCGCGCGCG 732  
Db 580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
Qy 733 GTGAAGCAGTGGCCCTGACCGGAGGAGAAGATCAAGGCCCTGACCGCCATCTCGGAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619  
Qy 793 ATGAGAGAGGAGGCAAGATCACCAAGATCGGCGCGCGAGAACCCCTACACACCGCGCTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
Qy 853 TTCGCCATGAAGAGGAGGACAGCACCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAAAGCGCACCGAGGACTTCTGGAGGTGCAGCTGGGCATCCCCACCGCGCGCGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAAGAGCGTGCAGCTGTGGAGCGCGTGCAGTGGCGGACCGCTACTTCAGCGTGCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGACTTCCGCAAGTACCGCCTTCACCATCCCGAGCATCAACAACGAGACCGCCC 1092  
Db 700 AspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
Qy 1093 GGCATCCGCTACCATGATCAACAGTCTGCCCCAGGCGCTGGAGGCGCGCGCGCATCTTC 1152  
|||||

Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
Qy 1153 CAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCTGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759  
Qy 1213 TACCAGTACATGAGCACCTGTACGTGGGCGAGCGACTCTGGAGATCGGCGAGCCGCGCC 1272  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1273 AGATCGAGAGCTGCGCAAGCACCTGTCTGCGCTGGGGTTCACACCCCGCGACAGAAG 1332  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
Qy 1333 CACCAAGAGGAGCGCCCTTCTGTGTGGTGGGTACGAGCTGCACCCCGACAGTGGACC 1392  
Db 800 HisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1393 GTGCAGCCCATCGAGCTGCGCGAAGAGGAGCTGGACCGTGAACGACATCCAGAAGCTG 1452  
Db 820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1453 GTGGGCAAGCTGAACCTGGGCGCGCAGATCTACCCCGCATCAAGGTGGCGCGCATGTGC 1512  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuCys 859  
Qy 1513 AAGCTCTGCGCGCGCGCAAGGCCCTGACCGCATCGTCCCTGACCGAGGAGGCGCGAG 1572  
Db 860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
Qy 1573 CTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGGAGCCCGTGCACCGCGTGTACTACGAC 1632  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899  
Qy 1633 CCAGCAAGAGACTGTGTGGCGGAGATCCAGAAGCAGGCGCGCACAGTGGAGCTACCAG 1692  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1693 ATCTACCGAGGCGCTTCAAGACCTGAAGACCGCAAGTACCGCAAGATGCGCACCGCC 1752  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1753 CACCAACAGCAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAGATCGCCATCGAGAGCATC 1812  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
Qy 1813 GTGATCTGGGCGAGACCCCGCAAGTTCGCGCTCCCATCCAGAGGAGAGCTGGGAGACC 1872  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1873 TGTGTGACCGACTACTGCGAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1932  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1933 CCCTCGTGAAGCTGTGTACCAAGCTGGAGAAGGAGCCCATCATCGCGCGCGAGACTTC 1992  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
Qy 1993 TAGCTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCGCTACGTGACCGAC 2052  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2053 CGGGCGCGCAGAAAGATCTGTAGCTGACCGAGACCAACCAACAGAAAGACCGAGCTGCAG 2112  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
Qy 2113 GCCATCCAGCTGGCCCTTCAGGAGCAGCGCGAGGTGAACATCTGTGACCGCAGCCAG 2172  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy 2173 TAGCGCTGGCGCATCATCCAGGCGCGCGAGCGAGAGGAGCGAGCTGGTGAACCGAG 2232  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluLeuValAsnGln 1099  
|||||

```
Qy 2233 ATCATCGAGCGAGTGTATCAAGAGAGAGAGTGTACCTGAGTGGTGGCGCCGACCAAG 2292
Db 1100 lfeileGluInLeulleLysGluLysValtyrLeuSerTrpValProAlaHisLys 1119
Qy 2293 GGCATCGCGCGCAAGCAGCATCGACAACTGGTGAGCAAGGCGCATCCGCAAGTGCTG 2352
Db 1120 GlylleGlyLysAenGluInleasplysleuValSerLysGlylleArgLysValleu 1139
Qy 2353 TTCTCGAGCGCATCGAT 2370
Db 1140 PheLeuaspGlylleasp 1145
RESULT 3
Q9WF86
ID Q9WF86 PRELIMINARY; PRT; 1428 AA.
AC
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polypeptide.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=9214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.;
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana."
RL J. Virol. 73:4427-4432(1999).
CC -!- SIMILARITY: Belongs to peptidase family A2.
DR EMBL; AF110973; AADI7119.1; -.
DR HSP; P04585; 1JUA.
DR GO; GO:0019012; C:varion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0015074; P:DNA integration; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn_N.
DR InterPro; IPR001995; Peptidase_A2.
DR InterPro; IPR009007; Pept_Aspatic.
DR InterPro; IPR001969; Pept_Asp_AS.
DR InterPro; IPR010999; Retrovir_matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR004477; RVTse.
DR InterPro; IPR010659; RVT-connect.
DR InterPro; IPR010661; RVT_thumb.
DR InterPro; IPR001878; Znf_CCHC.
Pfam; PF00540; Gag_p17; 1.
Pfam; PF00607; Gag_p24; 1.
Pfam; PF02022; Integrase_Zn; 1.
DR InterPro; IPR000075; RNaseH; 1.
```

```
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT_1; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF06817; RVT_thumb; 1.
DR Pfam; PF00098; ZF_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIVMATRIX.
DR SMART; SMO0343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
DR PROSITE; PS0158; ZF_CCHC; 2.
DR AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1428 AA; 161328 MW; 2D61CD9FEABD2387 CRC64;

Alignment Scores:
Pred. No.: 5,59e-159 Length: 1428
Score: 4170.00 Matches: 783
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 99.62% Mismatches: 0
Query Match: 90.99% Indels: 0
DB: 2 Gaps: 0

US-09-610-313B-30 (1-2469) x Q9WF86 (1-1428)
Qy 13 ATGGCCGAGGCGCATGAGCGCAGCCACAGCCCAACATCTGTGTCGAGCGAGCAATTC 72
Db 360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAlaLeuMetGlnArgSerAsnPhe 379
Qy 73 AAGGGCCCCAAGCCATCATCAAGTGTTCACCTGCGCAAGAGGGGCCACATCCCGCGC 132
Db 380 LysGlyProLysArgIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399
Qy 133 AACTGCCGCGCCCGCCGACAGAGGCTGCTGGAGTGGCGCAGAGGGCCACCAGATG 192
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419
Qy 193 AAGGACTGCACCGAGCGCCAGCCCACTTCTCCGCGAGGACCTGGCGCTTCCCCAGAGC 252
Db 420 LysAspCysThrGluArgGlnAlaAsnPheArgGluAspLeuAlaPheProGlnGly 439
Qy 253 AAGCGCCGCGAGTTCCCGAGCAGCAGAACCGCCCAACAGCCCGCCGAGCGAGCTG 312
Db 440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459
Qy 313 CAGTGGCGCGGACAAACCCCGCAGGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db 460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479
Qy 373 TTCCCGCAGATCACCTGTGTGCGAGCGCCCTGCTGAGCATCAAGTGGCGCGCGCGCGCG 432
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499
Qy 433 AAGGAGGCCCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519
Qy 493 GGCAGTGGGAAGCCCAAGATGATCGCGGCGCATCGCGCGCTTCATCAAGTGGCGCGCG 552
Db 520 GlyLysTrpLysProLysMetMetGlyGlyIleGlyPheIleLysValArgGlnTrp 539
Qy 553 GACCAGATCTGTATCGAGATCTGGCGGAAGAGCCCATCGGCACCGTGTCTCATCGGCC 612
Db 540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559
Qy 613 ACCCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACATTC 672
Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579
Qy 673 CCNATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGAGCGGCCCAAG 732
```



580	ProIleSerProIleGluThrValProValIysLeuIysProGlyMetAspGlyProLys	599
733	GTCAAGCAGTGGCCCTTACCGAGGAGAAGATCAAGGCCCTTGACCGCCATCTCGAGAG	792
600	ValIysGlnTrpProLeuThrGluGluIysIleIysAlaLeuThrAlaIleCysGluGlu	619
793	ATGAGAAGAGGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTWACAACACCCCGGTG	852
620	MetGluIysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal	639
853	TTCCGCATCAAGAGAGGACAGCAAGTGGCGCAAGTGGTGGAGCTTCGGAGCTG	912
640	PheAlaIleIysIysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
913	AACAAGCGCACCCAGGACTTCTGGAGGAGTGCAGCTGGGCATCCCCACCCCGCGGCTG	972
660	AsnIysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	679
973	AAGAAGAAGACGCTGACGCTGCGAGCGTGGCGACGCCCTACTTCAGCTGCCCTG	1032
680	LysLysLysIysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
1033	GACGAGACTTCGCAAGTACACGCCCTTACCACATCCCGAGCATCAACAACGAGACCC	1092
700	AspGluAspPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrPro	719
1093	GGCATCCGCTACCAAGTCAACGCTGCTGCCCGAGGGCTGGAAGGCGACCCAGCATCTTC	1152
720	GlyIleArgTrpGlnTyrAsnValLeuProGlnGlyTrpIysGlySerProSerIlePhe	739
1153	CAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATC	1212
740	GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle	759
1213	TACCAGTACATGACGACCTGTACTGCTGGGACGACCTGGAGATCGGCCACGACCGGCC	1272
760	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
1273	AAGATCCAGGAGCTGCCACGACCTGCTGGCTGGGGCTTCACCAACCCCGACACAGAAG	1332
780	LysIleGluLeuLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys	799
1333	CACCAGAAGGAGCCCTTCTGTGTGATGGGCTTACGAGCTGCACCCCGCAAGTGGAC	1392
800	HisGlnIysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	819
1393	GTGAGGCCCATCTGAGCTGCCGAGAGAGGAGCTGGACCGGTGAACGACATCCAGAGCTG	1452
820	ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu	839
1453	GTGGGCAAGCTGAAGTGGCCAGCCAGATCTACCCCGGCATCAAGGTGCCAGCTGTGC	1512
840	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleIysValArgGlnLeuCys	859
1513	AAGTGTGTGGCGCGCCAAAGGCCCTGACCGACATCTGTGCCCTTGACCGAGGAGCGCAG	1572
860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu	879
1573	CTGAGCTGGCCGAGAACCGCGAGATCCTGCGGAGCCCGGTGACGCGGTGTACTACGAC	1632
880	LeuGluLeuAlaGluAspArgIleLeuArgGluProValHisGlyValTyrTyrAsp	899
1633	CCGAGCAAGNACTGTGTGGCGGAGATCCAGAACGAGGGCCACGACCTGACCTACCAG	1692
900	ProSerLysAspLeuValAlaGluIleGlnIysGlnGlyHisAspGlnTrpThrTyrGln	919
1693	ATCTTACCAGAGCCCTTCAAGAACCTCAAGACCGGCAAGTAGCGCAAGATGCGCACCGCC	1752
920	IleTyrGlnGluProPheLysAsnLeuIysThrGlyLysTyrAlaLysMetArgThrAla	939
1753	CACACCAACGAGCTGAAGACGCTGACCGAGCCGTGCAGAAGATCGCCATCGGAGAGCATC	1812
940	HisThrAsnAspValIysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle	959

QY	1813	GTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGGGAGACC	18172
Db	960	ValliletrpGlylysthrPrProLysPheArgLeuProilleGlnlysthrTrpGluThr	979
QY	1873	TGGTGGACCCACTACTCTGGCAGGCCACTCGATCCCGAGTGGGAGTTCGTGAACACCCCC	1932
Db	980	TrpTrpThrAspTyTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	999
QY	1933	CCCTGTGTGAAGCTGTGGTACCGAGTGGAGAGAGCCCATCATCGCGCGCGAGACCTTC	1992
Db	1000	ProLeuValLysLeuThrTrpGlnLeuGluLysGluProillelleGlyAlaGluThrPhe	1019
QY	1993	TACGTGACCGCGCCGCAACCCCGAGACCAAGATCGGCAAGCCGCTACTCGTGAACCGAC	2052
Db	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysilleGlyLysAlaGlyTyValThrAsp	1039
QY	2053	CGGGCCCGCAGAAAGATCGTGAGCTTACCGAGACCAACCAACAGAGACCGAGCTGCAG	2112
Db	1040	ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln	1059
QY	2113	GCCATTCAGCTGGCCCTCGCAGGACAGCGGCGAGCGAGTGAACATCGTGACCGCAGCGCAG	2172
Db	1060	AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln	1079
QY	2173	TACGCCCTGGGCATCATCCAGGGCCCGCAGCAAGACGAGCGAGCGAGCTGGTGAACCGAC	2232
Db	1080	TyrAlaLeuGlyIlelleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln	1099
QY	2233	ATCATCGAGCAGCTGATCAAGAGGAGAGAGGTGTACTGAGCTGGTGGTCCGCCACCAAG	2292
Db	1100	IleIleGluGlnLeuIleLysLysGluLysValTyrlLeuSerTrpValProAlaHisLys	1119
QY	2293	GGCATCGCGCGCACGAGCAGATCGACAACCTGGTGAGCAAGGSCATCCCAAGGTGCTG	2352
Db	1120	GlyIleGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu	1139
QY	2353	TTCCTGCGCGCATCGAT	2370
Db	1140	PheLeuAspGlyIleAsp	1145
RESULT 4			
ID	Q9WF62	PRELIMINARY;	PRT; 1427 AA.
AC	Q9WF62;		
DT	01-NOV-1999	(TrEMBLrel. 12, Created)	
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)	
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)	
DE	Gag-pol polyprotein.		
OS	Human immunodeficiency virus 1.		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
OX	NCBI_TaxID=11676;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99214383; PubMed=10196340;		
RA	Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,		
RA	Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,		
RA	Essex M.;		
RT	"Molecular cloning and phylogenetic analysis of human immunodeficiency		
RT	virus type 1 subtype C: a set of 23 full-length clones from		
RT	Botswana.";		
RL	J. Virol. 73:427-4432(1999).		
CC	-I- SIMILARITY: Belongs to peptidase family A2.		
DR	EMBL; AF10961; AAD17021.1; -.		
DR	HSSP; Q70622; 1HVN.		
DR	GO; GO:0019012; C:viron; IEA.		
DR	GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.		
DR	GO; GO:0003677; F:DNA binding; IEA.		
DR	GO; GO:0008907; F:integrase activity; IEA.		
DR	GO; GO:0008233; F:ribonuclease H activity; IEA.		
DR	GO; GO:0004523; F:ribonuclease H activity; IEA.		
DR	GO; GO:0003723; F:RNA binding; IEA.		
DR	GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.		



```
QY 1393 GTGAGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTG 1452
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 820 ValGlnProIleuLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839
QY 1453 GTGGCAAGCTGAACCTGGGCCAGCAGCATATACCCCGCATCAAGGTGCGCCAGCTGTGC 1512
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyProGlyIleLysValArgGlnLeuCys 859
QY 1513 AAGCTGTGGCGGCCCAAGGCCCTGACGAGATCGTGCCCTGACCGAGAGGCGCGAG 1572
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879
QY 1573 CTGGAGCTGGCCGAGAACCGGAGATCTCGGGGAGCCCTGCGACGCGGTACTAGCAC 1632
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyTrpAsp 899
QY 1633 CCAGCAAGACCTGTGTGGCGGAGATCCAGAAGCAGGCCACACGACGACCTACCCAG 1692
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyGln 919
QY 1693 ATCTACGAGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCCAAGATGCGGACCGCC 1752
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 920 IleTyrglnGluProPheLysAsnLeuLysThrGlyLysTyrsAlaLysMetArgThrAla 939
QY 1753 CACACCAACGACGTGAGCAGCTCACCGAGCGCGCTGCAGAGATCGCCATGGAGAGCATC 1812
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle 959
QY 1813 GTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACC 1872
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979
QY 1873 TGTGGACCGACACTACTGGCAGGCCACTGTGATCCCGAGTGGAGTTCGTGAACACCCCC 1932
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 980 TrpTrpThrAspTyTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999
QY 1933 CCCTGTGTGAGCTGTGTACCGAGTGGAGAGGAGCCCATCATCGCGCGGAGACCTTC 1992
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1000 ProLeuValLysLeuTrpTyrglnLeuGluGluGluProIleValGlyAlaGluThrPhe 1019
QY 1993 TAGCTGGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2052
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrsValThrAsp 1039
QY 2053 CGGGCGCGGAGAGATCGTGTAGCTGACCGAGACACCAACAGAGACCGAGCTGCAG 2112
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1040 ArgGlyArgGlnArgIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059
QY 2113 GCCATCCAGCTGGCCCTGCAGGACGCGGAGCGAGTGAACATCGTGACCGACGACGAG 2172
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln 1079
QY 2173 TAGCCCTGGCGCATCATCCAGCGCCAGCCGACAGAGAGCGAGCGAGCTGGTGAACACGAG 2232
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099
QY 2233 ATCATCGAGCAGCTGTATCAAGAGGAGAGGTGTACTGTAGCTGGGTGCCCGCCACAG 2292
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1100 IleIleGluGlnLeuIleLysLysGluArgValTyrsLeuSerTrpValProAlaHisLys 1119
QY 2293 GGATCGGGCGCAACGAGCAGATCGACAAGCTGTGTAGCAAGGCGCATCCGCAAGGTGCTG 2352
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1120 GlyIleGlyLysGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139
QY 2353 TTCTTGGACGGCATCGAT 2370
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1140 PheLeuAspGlyIleAsp 1145
RESULT 5
Q6X4P8 PRELIMINARY; PRT; 1429 AA.
ID Q6X4P8
AC Q6X4P8;
```

```
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Gag-poi fusion polyprotein (Fragment).
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14709249; DOI=10.1089/0892220371881220;
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,
RA McCutchan F.E.;
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT in full genome analysis, is not sustained in subgenomic regions.";
RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Harris M.E., Birx D.L.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kim B., McCutchan F.E.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Maayan S., Zeira M.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family A2.
DR HSPB; P03367; 1A8G.
DR GO: GO:0019012; C:virion; IEA.
DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0008907; F:integrase activity; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004523; F:ribonuclease H activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0015074; P:DNA integration; IEA.
DR GO: GO:0006310; P:DNA recombination; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO: GO:0016032; P:viral life cycle; IEA.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_Zn_N.
DR InterPro: IPR001995; Peptidase_A2.
DR InterPro: IPR009007; Pept_Aspatic.
DR InterPro: IPR001969; Pept_Asp_AS.
DR InterPro: IPR010999; Retrovir_matrix.
DR InterPro: IPR000071; Retrovir_p17.
DR InterPro: IPR008916; Retrov_capsid_C.
DR InterPro: IPR002156; Retrov_capsid_N.
DR InterPro: IPR001584; RNaseH.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR010659; RVT_connect.
DR InterPro: IPR010661; RVT_thumb.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00540; Gag_p17; 1.
DR Pfam: PF00607; Gag_p24; 1.
DR Pfam: PF02022; Integrase_Zn; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00077; RVP; 1.
DR Pfam: PF00078; RVT_1; 1.
DR Pfam: PF06815; RVT_connect; 1.
DR Pfam: PF06817; RVT_thumb; 1.
DR Pfam: PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
```





Db 418 LysAspCysThrGluArgGlnAlaAsnPheArgGluIleLeuAlaPheProGlnGly 437  
Qy 253 AAGCCCGCGAGTTCCTCCAGGAGCAGAACCGCGCCCAACAGCCCAACAGCGCGAGCTG 312  
Db 438 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrAsnGlyGluLeu 457  
Qy 313 CAGGTGGCGCGGAGCAACACCCCGAGGAGCCCGCGCGAGCGCGAGCGGCGACCTGAAC 372  
Db 458 GlnValArgArgAsnAsnProArgSerLysThrGlyValGluArgGlnGlyThrLeuAsn 477  
Qy 373 TTCCCCAGATCACCTGTGTGACGCGCCCTGTGTGAGCATCAAGSTGGGGCGGCAGATC 432  
Db 478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 497  
Qy 433 AAGGAGCCCTGTGTGACACCGCGCGCGAGACACCGTGTGTGAGAGAGAGCTGCCCC 492  
Db 498 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517  
Qy 493 GGCAGGTGGAGAACCAAGATGATCGGCGGCATCGCGCGCTTCATCAAGGTGCGCGCATAC 552  
Db 518 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 537  
Qy 553 GACGAGATCTGTGAGATCTGGGCAAGAGCCATCGGACCGTGTGATCGGCGCC 612  
Db 538 AspGlnIleLeuIleGluIleCysGlyLysAlaIleGlyThrValLeuValGlyPro 557  
Qy 613 ACCCCCTGAAACATCATCGGCGCAACATGTGTGACCGAGCTGGCGTGCACCTGAACCTC 672  
Db 558 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 577  
Qy 673 CCATACAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGGCCCAAG 732  
Db 578 ProfileSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597  
Qy 733 GTGAAGCAGTGGCCCTGACCGAGGAGNATCAAGCCCTGACCGGCATCTGCGAGGAG 792  
Db 598 ValLysGlnTrpProLeuThrGluGluLysAlaLeuThrAlaIleCysGluGlu 617  
Qy 793 ATGAGAAGAGGAGGCAAGATCACCAAGATCGGCCCGAGAACCCCTCAACACACCCCGTG 852  
Db 618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 637  
Qy 853 TTGCGCATCAAGAAGAAGAGCAGCACCAAGTGGCGCAAGCTGTGAGCTTCGCGAGCTG 912  
Db 638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657  
Qy 913 AACAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGCGCATCCCGCACCCCGCGGCTG 972  
Db 658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677  
Qy 973 AAGAAGAAGAAGAGCGTGTGACGCTGAGCGTGGCGGACGCTACTTTCAGCGTCCCGCTG 1032  
Db 678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 697  
Qy 1033 GACGAGGACTTCGCAAGTACACCGCTTCACCATCCCGAGCATCAACACAGAGACCCCC 1092  
Db 698 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717  
Qy 1093 GGCATCCGCTACCAAGTACAGCTGCTCGCCAGGCTGGAGGGCAGCCCGCAGCATCTTC 1152  
Db 718 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 737  
Qy 1153 CAGAGCAGCATGACCAAGATCTCGAGCCCTTCGCGCGCGCAACCCCGAGATCGTGTATC 1212  
Db 738 GlnSerSerMetThrArgIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 757  
Qy 1213 TACCAGTACATGACGACCTGTATCGTGGGAGCGACCTGGAGATCGCGCCAGCACCGCGCC 1272  
Db 758 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 777  
Qy 1273 AAGATCAGGAGCTGCGCAAGCACCTGCTGGCTGGGGCTTCACCAACCCCGCAAGAG 1332

Db 778 LysIleGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797  
Qy 1333 CACCAGAAGAGCCCCCTTCTCTGTGATGGCTACTCAGAGCTGCACCCCGCAAGTGGACC 1392  
Db 798 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 817  
Qy 1393 GTGCAGCCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGATCATCCAGAGCTG 1452  
Db 818 ValGlnProIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 837  
Qy 1453 GTGGCGAAGCTGAATCGGCGCAGCAGATCTACCCCGGCATCAAGGTGCGCAGCTGTC 1512  
Db 838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 857  
Qy 1513 AAGTGTGTGCGCGCCCAAGCCCTGACCGACATCGTGCCTCCCTGACCGAGAGCCGAG 1572  
Db 858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 877  
Qy 1573 CTGGAGCTGCGCGAGAACCCGCGAGATCTCTGCGCGAGCCCTGTCACGGGCTGTACTACGAC 1632  
Db 878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp 897  
Qy 1633 CACGAGCAAGCAGCTGTGGCGCGAGATCCAGAAAGCAGGCGCACCGACCTAGTGCAC 1692  
Db 898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnLysAspAspGlnTrpThrTyrGln 917  
Qy 1693 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCCGCGCAAGTACGCGCAAGATCGCACCGCC 1752  
Db 918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937  
Qy 1753 CACACCAACGAGTGAAGAGCTGACCGAGGCGCTGCGAAGATCGCCATCGGAGAGCATC 1812  
Db 938 HisThrAsnAspValLysGlnLeuValGlyAlaValGlnLysIleAlaMetGluGlyIle 957  
Qy 1813 GTGATCTGGCGGAGAGCCCGCAAGTCTCGCTGCCCATCCAGAAAGGAGACCTGGGAGACC 1872  
Db 958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 977  
Qy 1873 TGGTGGACCGACTACTGCGAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1932  
Db 978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPheValAsnThrPro 997  
Qy 1933 CCGCTGTGAAGCTGTGGTACCAGCTGAGAGAGGCCCATCATCTCGGCGCGAGACCTTC 1992  
Db 998 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyValGluThrPhe 1017  
Qy 1993 TAGCTGAGCGCGCGCCCAACCGGAGACCAAGATCGCAAGCGCGCTACTGTCACCGAC 2052  
Db 1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037  
Qy 2053 CGGCGCGCGCAGAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2112  
Db 1038 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057  
Qy 2113 GCCATCAGCTGCGCCCTGCGAGGACGCGGCGAGGTGAACATCTGTGACCGAGCGCAG 2172  
Db 1058 AlaIleGlnIleAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1077  
Qy 2173 TAGCCCTGGGCATCATCCAGGCGCCAGCCGACAAAGAGCGAGCGAGCTGGTCAACCCAG 2232  
Db 1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097  
Qy 2233 ATCATCCAGCAGCTGTATCAAGAAGGAGAGTGTATCTGAGCTGGGTGCCCGCCCAAG 2292  
Db 1098 IleIleGluGlnLeuIleLysGluArgIleTyrLeuThrTrpValProAlaHisLys 1117  
Qy 2293 GGCATCGCGCGCACGAGCAGATCGAACAGCTGGTGAAGGCGCATCCCGCAAGGTGCTG 2352  
Db 1118 GlyIleGlyAsnGluGlnValAspLysLeuValSerAsnGlyIleArgLysValLeu 1137  
Qy 2353 TTCTCGACCGCATCGAT 2370  
Db 1138 PheLeuAspGlyIleAsp 1143





Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAGCGCACCCAGGACTTCTGGAGAGTGGAGCTGGGCATCCCGACCCCGCGGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAAGAAGCGTGACCGTCTGTGAGCGTGGCGACGGCTACTTTCAGCGTGCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGGACTTCCCAAGTACACCGCTTCCACCATCCCGACCATCAACAAGAGACCC 1092  
Db 700 AspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719  
Qy 1093 GGCACTCCCTACCACTACAACTGCTGCTCCCGAGGCTGGAGGGCAGCCCGACGATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
Qy 1153 CAGAGCATGATGACCAAGATCCTGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759  
Qy 1213 TACCAGTACATGAGCAGCTGTAGTGGGAGCGACCTGGAGATCGGCGCACCGCGCC 1272  
Db 760 TyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1273 AAGATCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCCACCCCGCAAGAAG 1332  
Db 780 LysIleGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799  
Qy 1333 CACCAAGAGAGCCCTTCTGTGATGGCTACGAGCTGCACCCCGCAAGTGGACC 1392  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1393 GTGCAGCCCTCGAGCTGCCGAGAGGAGCTGGACCTGGACCATCCCAAGCTG 1452  
Db 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1453 GTGGCAAGCTGAACTGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTC 1512  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
Qy 1513 AAGTGTGCGCGCGCCAGGCTCTGACGACATCTGTGCTCCCTGACGAGAGCGCAG 1572  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879  
Qy 1573 CTGAGCTGGCGAGACCGAGATCTGCGCAGCGCTGACGCGCTGTACTACGAC 1632  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLysLysGluProValHisGlyValTyrTyrAsp 899  
Qy 1633 CCCAGCAGGACCTGTGGCGCAGATCCAGAGCGGCGCACGACCATGAGCTACACG 1692  
Db 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1693 ATTCACGAGGCTCTTCAAGAACCTGAGACCGCGCAAGTACGCGCAAGATCGCGACCGCC 1752  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1753 CACACCAACGCTGAGCAGCTGACCGAGCGCTGCAGAGATCGCCATCGAGCAGCATC 1812  
Db 940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959  
Qy 1813 GTGATCTGGGCAAGACCCCAAGTCTCGCTCGCCATCCAGAGGAGAGCTGGGAGACC 1872  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1873 TGTGGACCGCACTACTGGCAGGCCCTTGGATCCCGAGTGGGAGTTCGTGAACACCC 1932  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1933 CCGCTGTGAGCTGTGGTACAGCTGAGAGAGCCCATCATCGCGCGCGAGACCTTC 1992  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy 1993 TACGTGACGCGCGCCGCAACCGCGAGACCAAGATCGCAAGCGCGGTACGTGACCCAC 2052  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2053 CGGGCGCGCAGAGATCGTGAGCTGACCGCGAGACCAACCAACAGACCGAGCTGCAG 2112  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
Qy 2113 GCCATCCAGCTGGCGCTGCGAGCAGCGCGAGGAGGTGAACATCGTGACCGACGAC 2172  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079  
Qy 2173 TACGCTCTGGCATCATCCAGCGCCCGCAAGAGCGAGCGAGCTGCTGTAACCCAG 2232  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
Qy 2233 ATCATCCAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTGGCGCCGACCAAG 2292  
Db 1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
Qy 2293 GGCATCCGCGCGCAACGAGCAGATCGCAAGCTGTGAGCGAGCGCATCCGCAAGTGTG 2352  
Db 1120 GlyIleGlyGlnAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
Qy 2353 TTCCTGACGCGCATCGAT 2370  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 8  
Q6X4R4 ID Q6X4R4 PRELIMINARY; PRT; 1430 AA.  
AC Q6X4R4;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE gag-pol fusion polyprotein (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14709249; DOI=10.1089/089922203771881220;  
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,  
RA McCutchan F.E.;  
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
in full genome analysis, is not sustained in subgenomic regions.";  
RL AIDS Res. Hum. Retroviruses 19:1125-1133 (2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Harris M.E., Birx D.L.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kim B., McCutchan F.E.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Maayan S., Zeira M.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AY255823; AAP76521.1; -;  
DR HSP; P12497; I89D.  
DR GO; GO:0019012; C:viron; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:phosphatase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003664; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.

DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0015074; P:DNA integration; IEA.  
DR GO: GO:0006310; P:DNA recombination; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO: GO:0016032; P:viral life cycle; IEA.  
DR InterPro: IPR000721; Gag\_p24.  
DR InterPro: IPR001037; Integrase\_C.  
DR InterPro: IPR003308; Integrase\_Zn\_N.  
DR InterPro: IPR001995; Peptidase\_A2.  
DR InterPro: IPR009007; Peptidase\_A2.  
DR InterPro: IPR001969; Peptidase\_A2.  
DR InterPro: IPR010999; Retroviral matrix.  
DR InterPro: IPR000771; Retroviral p17.  
DR InterPro: IPR008916; Retroviral capsid\_C.  
DR InterPro: IPR008919; Retroviral capsid\_N.  
DR InterPro: IPR002156; RNaseH.  
DR InterPro: IPR001584; Rve.  
DR InterPro: IPR000477; RVTse.  
DR InterPro: IPR010659; RVT connect.  
DR InterPro: IPR010661; RVT\_thumb.  
DR InterPro: IPR001878; Znf\_CCHC.  
DR Pfam: PF00540; Gag\_p17; 1.  
DR Pfam: PF00607; Gag\_p24; 1.  
DR Pfam: PF02022; Integrase\_Zn; 1.  
DR Pfam: PF00075; RNaseH; 1.  
DR Pfam: PF00665; Rve; 1.  
DR Pfam: PF00077; RVP; 1.  
DR Pfam: PF00078; RVT\_1; 1.  
DR Pfam: PF06815; RVT connect; 1.  
DR Pfam: PF06817; RVT\_thumb; 1.  
DR Pfam: PF00098; zf\_CCHC; 2.  
DR PRINTS: PR00939; C2HCZNFINGER.  
DR SMART: SM00234; HIVMATRIX.  
DR SMART: SM00343; Znf\_C2HC; 2.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
DR PROSITE: PS01175; ASP\_PROT\_RETROV; 1.  
DR PROSITE: PS01158; ZF\_CCHC; 2.  
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
KW Zinc-finger.  
FT NON TER  
SQ SEQUENCE 1430 AA; 161540 MW; 5BBA39B85D03FA0B CRC64;

Alignment Scores:  
Pred. No.: 1,81e-151 Length: 1430  
Score: 3982.00 Matches: 735  
Percent Similarity: 97.59% Conservative: 36  
Best Local Similarity: 93.04% Mismatches: 15  
Query Match: 86.89% Indels: 4  
DB: 2 Gaps: 1

US-09-610-313B-30 (1-2469) x Q6X4R4 (1-1430)

QY 313 CAGGTGCGGGGCAACACCCCGCAGGCGCGCGCGCGC-----CAG 360  
DB 459 GlnValArgGlyAspAsnProLeuSerGluAlaGlyThrGluArgProGlyThrLeuGln 478  
QY 361 GGACCCCTGAATTCCTCCAGATCACCTGTGTGGACGGCCCTGTGTGAGCATCAAGGTG 420  
DB 479 GlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLeu 498  
QY 421 GGCGGCAGATCAAGGAGCGCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
DB 499 GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGlu 518  
QY 481 ATGAGCTGCTCCCGCGCAAGTGGAGGCAAGATCATCGCGCGCATCGCGCGCTTCATCAAG 540  
DB 519 IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyLeuGlyPheLeuLys 538  
QY 541 GTGCGCCAGTACGACCAAGATCTGTGAGATCTGTGGCAAGAGCGCATCGGACCGGTG 600  
DB 539 ValArgGlnTrpAspGlnIleLeuIleGluLeuLeuLysGlyLysAlaIleGlySerVal 558  
QY 601 CTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCGCGCTGGCTGC 660  
DB 559 LeuValGlyProThrProValAsnIleGlyArgAsnMetLeuThrGlnLeuGlyCys 578  
QY 661 ACCTGAACTTCCCATCGACCCCATCGACCGCTGCTGCGTGAAGCTGAAGCGCGCGCATG 720  
DB 579 ThrLeuAsnPheProIleSerProIleGluThrIleProValLysLeuLysProGlyMet 598  
QY 721 GACGCGCCCAAGGTGAAGCATGTGGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCC 780  
DB 599 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAla 618  
QY 781 ATCTGCGAGGAGTGAAG 840  
DB 619 IleCysGluAspMetGluArgGluGlyLysIleSerLysIleGlyProGluAsnProTyr 638  
QY 841 AACACCCCGCTGTTCGCCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 639 AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp 658  
QY 901 TTCGCGAGCTGAACAAAGCGACCCAGGACTTTCGCGAGAGTGTGGAGTGTGGAGTGTGG 960  
DB 659 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 678  
QY 961 CCGCGCGCGCTGAAG 1020  
DB 679 ProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 698  
QY 1021 AGCGTGGCGCTGACGAGGAGACTTTCGCGAGTACACCGCGCTTCCACCATCCCGCAGCATCAAC 1080  
DB 699 SerValProLeuAspGluAsnPheArgLysTyrThrAlaPheThrIleProSerThrAsn 718  
QY 1081 AACGAGACCCCGCATTCGCTACAGTACAAAGTGTGCTGCGCGCGCGCGCGCGCGCGCGCG 1140  
DB 719 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyrLysGlySer 738  
QY 1141 CCGAGCATCTTCAGAGGAGCATGACCAAGATCCTGAGCGCTTCCGCGCGCGCGCAACCCC 1200  
DB 739 ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro 758  
QY 1201 GAGATCGTGTATCTACAGTACATGAGCGAGCTGTAGTGGCGAGCGACCTGGAGATCGGC 1260  
DB 759 GluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 778  
QY 1261 CAGACCGCGCCCAAGATCGAGGAGTGTGGCAAGCATCTGTGCTGGCTGGCGCTTCACACC 1320  
DB 779 GlnHisArgAlaLysValGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThr 798  
QY 1321 CCGCAG 1380  
DB 799 ProAspLysHisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisPro 818

Qy 1381 GACAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGCAGCGTGAACGAC 1440  
 Db AsplysitrprthrValGlnThrIleGlnLeuProGluLysAspSerTrpThrValAsnAsp 838  
 Qy 1441 ATCCAGAGCTGGTGGGCAAGCTGAACCTGAGCGCCAGCAGATCTACCCCGGATCAAGGTG 1500  
 Db IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 858  
 Qy 1501 CGCAGCTGTGCAAGCTGCTGGGGGCGCCCAAGGCCCTGACCGCATCGTCCCTGACGC 1560  
 Db ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValAlaLeuThr 878  
 Qy 1561 GAGGAGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCGCGGAGCCCGTGCAGCGC 1620  
 Db GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValIleGly 898  
 Qy 1621 GTGTACTACGACCCAGCAGGACTGGTGCCGAGATCCAGAGATCCAGAGCGGGCCAGCAG 1680  
 Db ValTyrTyrAspProSerLysAspLeuIleAlaGluLeuGlnLysGlnGlyHisAspGln 918  
 Qy 1681 TGGACCTACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740  
 Db TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 938  
 Qy 1741 ATGCGCACCGCCACACCAACGAGCTGAGAGCTGACGAGCGCGTGCAGAGATCGCC 1800  
 Db ArgArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 958  
 Qy 1801 ATGAGAGACCTCGAGTCTGGGGCAAGACCCCAAGTTCGSCCTGCCATCCAGAGGAG 1860  
 Db MetGluSerIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGlu 978  
 Qy 1861 ACCTGGGAGACCTGGTGAGACGACTACTTGGCAGGCGCCACCTGGATCCCGAGTGGGAGTTC 1920  
 Db ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 998  
 Qy 1921 GTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCATCATCGGC 1980  
 Db ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 1018  
 Qy 1981 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGC 2040  
 Db AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1038  
 Qy 2041 TACGTGACCGACCGCGCGCGCAGAGATCTGAGCCTGACCGAGACCCACCAACGAGAAG 2100  
 Db TyrValThrAspLysGlyArgGlnLysValValSerLeuAsnGluThrThrAsnGlnLys 1058  
 Qy 2101 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGCAGCAGCGCGCAGCGGAGTGAACATCGTG 2160  
 Db ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1078  
 Qy 2161 ACCGACAGCAGTACGCGCTGGGATCATCTACAGAGGAGAGGTGTACCTGAGCTGGGTG 2220  
 Db ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerAspSerGlu 1098  
 Qy 2221 CTGCTGAACAGATCATCGACGACTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2280  
 Db IleValAsnGlnIleIleGlnLeuIleLysLysGluArgValTyrLeuSerTrpVal 1118  
 Qy 2281 CCCGCCACAGGAGCATCGCGGCAACGAGCAGATCGACAGCTGGTGGAGCAAGCGGCATC 2340  
 Db ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerLysGlyIle 1138  
 Qy 2341 CGCAGAGTGTCTTCTGGAGCGGATCGAT 2370  
 Db ArgLysValLeuPheLeuAspGlyIleAsp 1148  
 RESULT 9  
 Q9WF71  
 ID Q9WF71 PRELIMINARY; PRT: 1427 AA.  
 AC Q9WF71;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RL J. Virol. 73:4427-4432(1999).  
 CC -1- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AF110964; AAD17047.1; -;  
 DR HSP; Q70622; 1HWV.  
 DR GO; GO:0019012; C:Viron; IEA.  
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008907; F:integrase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:00016740; F:structural molecule activity; IEA.  
 DR GO; GO:0006278; F:transferase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_P24.  
 DR InterPro; IPR001037; Integrase C.  
 DR InterPro; IPR003308; Integrase Zn N.  
 DR InterPro; IPR001995; Peptidase A2.  
 DR InterPro; IPR009007; Pept Aspartic.  
 DR InterPro; IPR001969; Pept\_Asp AS.  
 DR InterPro; IPR010999; Retrovirk\_matrix.  
 DR InterPro; IPR000071; Retrovir\_p17.  
 DR InterPro; IPR008916; Retrov\_capsid C.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR004477; RVTse.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF06815; RVT\_connect; 1.  
 DR Pfam; PF06817; RVT\_thumb; 1.  
 DR Pfam; PF00098; zf\_CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PRINTS; PR00234; HIV1MATRIX.  
 DR SMART; SM00343; Znf\_C2HC; 2.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS00158; ZF\_CCHC; 2.  
 DR AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
 KW Zinc-finger.  
 SQ SEQUENCE 1427 AA; 161326 MW; B9FC43C29B26CFA2 CRC64;

Alignment Scores:

Pred. No.:	3,96e-151	Length:	1427	QY	973	AAGAGAGAGAGAGCGCTGACCGCTGTGGAGCGCGCTTACTTACGCGTGCCTG	1032
Score:	3973.50	Matches:	731	Db	680	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
Percent Similarity:	97.46%	Conservative:	35	QY	1033	GACGAGGACTTCGCGAAGTACACCGCTTCAACATCCAGCATCAACACGAGACCC	1092
Best Local Similarity:	93.00%	Mismatches:	19	Db	700	AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro	719
Query Match:	86.70%	Indels:	1	QY	1093	GGCATCCGCTACAGTACACGCTGCTGCCCGCTGAAGCGGAGCGCCAGCATCTTC	1152
DB:	2	Gaps:	1	Db	720	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyrPheSerProAlaIlePhe	739
US-09-610-313B-30 (1-2469) x Q9WF71 (1-1427)				QY	1153	CAGAGCAGCATGACCAAGATCTCTGGAGCCCTCCGCGCCGCGCAACCCCGAGATCGT	1212
				Db	740	GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle	759
				QY	1213	TACCATGATGAGCAGCTGTACGTGGGAGGAGCTGGAGATCGGCGAGACCGCGCC	1272
				Db	760	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
				QY	1273	AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTACACACCCCGCAAGAAG	1332
				Db	780	LysIleGluGluLeuArgGlnHisLeuLeuLysTyrGlyLeuThrThrProAspLysLys	799
				QY	1333	CACCAGAGGAGCCCTTCTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGACC	1392
				Db	800	HisGlnLysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyrThr	819
				QY	1393	GTGACGCCATCGAGCTGCGCGAGAGGAGAGCTGGACCGTGAACGATCCGAGAGCTG	1452
				Db	820	ValGlnProIleGlnLeuProAspLysAspSerTyrThrValAsnAspIleGlnLysLeu	839
				QY	1453	GTGGGCAAGCTGAACCTGGCGCAGCATCTACCCCGCATCAAGGTGGCCAGCTGTGC	1512
				Db	840	IleGlyLysLeuAsnTyrPheGlnIleTyrProGlyIleArgValLysHisLeuLys	859
				QY	1513	AAGCTCTGCGCGCGCCCAAGCCCTGACCGACATCGTCCCTGACCGAGGAGCGGAG	1572
				Db	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu	879
				QY	1573	CTGGAGCTGGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACCGCTGTACTAGC	1632
				Db	880	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	899
				QY	1633	CCAGCAAGACCTGTGGCGGAGATCCAGAACGAGCGCCACACAGTGCACCTACCG	1692
				Db	900	ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTyrThrGln	919
				QY	1693	ATCTACCGAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCCAAGATGCGCAC	1752
				Db	920	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
				QY	1753	CACACCAAGCATGAGCGAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATC	1812
				Db	940	HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleAlaMetGluCysIle	959
				QY	1813	GTGATCTGGCGCAAGACCCCAAGTTCCGCTCCCATCCAGAACGAGACCTGGGAGACC	1872
				Db	960	ValIleTyrLysThrProLysPheArgLeuProIleGlnLysAspThrTyrGluThr	979
				QY	1873	TGGTGGACCGACTACTGGCAGGCCACTGGATCCCGAGTGGGAGTTCGTGAACACCC	1932
				Db	980	TyrThrAspTyrTyrGlnAlaThrTyrIleProGluTyrGluPheValAsnThrPro	999
				QY	1933	CCCTGTGTGAGCTGTGTACAGCTGGAGAGGCCCATCATCGCGCGGAGACCTTC	1992
				Db	1000	ProLeuValLysLeuTyrTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe	1019
				QY	1993	TAGTGTGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCG	2052
				Db	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp	1039
				QY	2053	CGGGCGCGCAGAGATCGTGAGCCTGACCGAGACCAACCAACAGAGACCGAGCTG	2112

```

Db 1040 LysGlyArgGlnLysIleValSerLeuAenGluThrThrAenGlnLysAlaGluLeuGln 1059
Qy 2113 GCCATCCAGCTGGCCCTCAGCAGCAGCGCAGCGAGGTGAACATCGTGACCGCAGCCAG 2172
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAenIleValThrAspSerGln 1079
Qy 2173 TACGCTGGCAGCATCATCCAGCCCGCAGCCGACAGAGCGAGCGAGCTGGTGAACCCAG 2232
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099
Qy 2233 ATCATCGAGCAGCTGATCAAGAAGCAGAGAGCTGACCTGAGCTGGTGGCCGCCACAG 2292
Db 1100 IleIleGluGlnLeuIleLysGlyGluArgValTyrLeuSerTrpValProAlaHisLys 1119
Qy 2293 GGCATCGCGCGCAACAGCAGCATCCACAAGCTGGTGAAGGCGATCCGCAAGGTGCTG 2352
Db 1120 GlyIleGlyAenGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139
Qy 2353 TTCCTGAGCGCATCGAT 2370
Db 1140 PheLeuAepGlyIleAsp 1145

RESULT 10
Q9WFA0 PRELIMINARY; PRT; 1427 AA.
AC Q9WFA0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OC NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana."
RL J. Virol. 73:4427-4432(1999).
CC -!- SIMILARITY: Belongs to peptidase family A2.
DR EMBL; AF110978; AAD17162.1; -.
DR PIR; S49086; S49086.
DR HSP; Q70622; 1HVN.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0015074; F:DNA integration; IEA.
DR GO; GO:0006310; F:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR007021; Gag_p24.
DR InterPro; IPR001037; Integrase C.
DR InterPro; IPR003308; Integrase Zn N.
DR InterPro; IPR001995; Peptidase A2.
DR InterPro; IPR009007; Pept Aspartic.
DR InterPro; IPR001969; Pept_Asp_AS.
DR InterPro; IPR010999; Retrovir_matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.

```

```

DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR010659; RVT_connect.
DR InterPro; IPR010661; RVT_thumb.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF02022; Integrase Zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; Rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT_1; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF06817; RVT_thumb; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIVMATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
DR PROSITE; PS0158; ZF_CCHC; 2.
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polyprotein; Protease; RNA-directed DNA polymerase; transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1427 AA; 160503 MW; 24C410797550C0F8 CRC64;

Alignment Scores:
Pred. No.: 6e-151 Length: 1427
Score: 3969.00 Matches: 731
Percent Similarity: 97.46% Conservative: 35
Best Local Similarity: 93.00% Mismatches: 20
Query Match: 86.60% Indels: 0
DB: 2 Gaps: 0

US-09-610-313B-30 (1-2469) x Q9WFA0 (1-1427)
Qy 13 ATGCCCGAGCGCATGAGCGAGCGCCAGCCAGCCCAACATCTCTGATGCGAGCGAGCAACTTC 72
Db 360 LeuAlaGluAlaMetSerGlnAlaAsnThrAsnIleMetIleGlnArgSerAsnPhe 379
Qy 73 AAGGCGCCCAAGCGCATCATCAAGCTGTTCACTGCGGCAAGAGGGCCACATCGCCCGC 132
Db 380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399
Qy 133 AACTGCGCGCGCCCGCCCAAGAGGGCTGCTGGAGTGCGCGAGGCGGCGCCAGAGCCAGATG 192
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419
Qy 193 AAGGACTGCGCCAGCGCGCCAGCCCAACTTCTTCCGCGAGGAGCTGCTTCCCGCAGGC 252
Db 420 LysAspCysThrGluArgGlnAlaAspPhePheArgGluAenLeuAlaPheProGlnGly 439
Qy 253 AAGCGCGCGAGTTCCCGCAGCGAGCAGAAACCGCGCCCAACAGCCCGCCAGCGCGAGCTG 312
Db 440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerGlyGluLeu 459
Qy 313 CAGGTGCGCGCGCAACACCCCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db 460 GlnValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479
Qy 373 TTCGCCCGAGATCACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnVal 499
Qy 433 AAGGAGCGCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519
Qy 493 GGCNAGTGGAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 552
Db 520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyr 539

```



```
QY 553 GACACAGTCTGTGATCGAGATCTCGGCAAGAAGCCCATCGGCACCGTGTGATCGGCCCC 612
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
540 AspGlnIleThrIleGluIleCysGlyIysIysAlaIleGlyThrValLeuValGlyPro 559
QY 613 ACCCCCGTGAACATCATCGCCGCAACATCTGACCCAGCTGGGCTGACACCTGAATTC 672
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe 579
QY 673 CCATCAGCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGCATGGACGCCCCCAAG 732
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599
QY 733 GTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCATCTCGAGGAG 792
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
600 ValIysGlnTrpProLeuThrGluIuIysIleIysAlaLeuThrAlaIleCysGluGlu 619
QY 793 ATGGAGAAGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACAAACACCCCGTG 852
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
620 MetGluLysGluGlyIysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639
QY 853 TTGCCCATCAAGAAGAGGACAGCACCAAGTGGCGGAGCTGGTGGACTTCCGCGAGCTG 912
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659
QY 913 AACAAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCACCCCGCGGCTG 972
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679
QY 973 AAGAAGAAGAGCGCTGACCGTGTGGAGCTGGGCGACCGCTACTTTCAGCGTGCCTG 1032
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699
QY 1033 GACGAGACTTCCGCAAGTACACCGCCTTCAACATCCCGAGCATCAACACGAGACCCCC 1092
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 AspGluSerPheArgLysIleThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719
QY 1093 GGATCCGCTACCAAGTACACGCTGCTCCCGAGGCTGGAAGGCGACGCCCGAGCATCTTC 1152
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739
QY 1153 CAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759
QY 1213 TACCAAGTACATGACGACCTGTAGCTGGGAGCGACCTGAGATCGCGCAGCACCGCGCC 1272
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779
QY 1273 AAGATCGAGGAGCTGCGCAAGCACCTCTGCTGGGCTTCAACACCCCGCAAGAAAG 1332
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
780 LysIleGluGluLeuArgThrHisLeuLeuIysTrpGlyPheThrThrProAspLysLys 799
QY 1333 CACCAAGAAGGAGCCCCCTTCTGTGATGGCTTACGAGCTGCACCCCGCAAGTGGACC 1392
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819
QY 1393 GTGACGCCATCGAGCTGCCGAGAGGAGAGTGGACCGTGAACGACATCCAGAGCTG 1452
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839
QY 1453 GTGGCAAGCTGAACCTGGGCCCGCACATCTACCCCGGCATCAAGGTGGCCGAGCTGTGC 1512
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859
QY 1513 AAGTGTCTGGCGCGCCCAAGGCCCTGACCGACATCGTGCCTCGCCCTGACCGAGGAGCCGAG 1572
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879
QY 1573 CTGAGCTGGCCGAGAACCCGCGAGATCTCGCGGAGCCCGTGCACGCGCGTGTACTAGCAC 1632
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp 899
```

```
QY 1633 CCACCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGCGCACCGACCTGAGACTACCAG 1692
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919
QY 1693 ATCTACAGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACCGCAAGATGCCACCGCC 1752
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939
QY 1753 CACCAACACGACGTGAAGCAGCTGACCGAGGCGCTGACAGAAGATCGCCATCGAGAGCATC 1812
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959
QY 1813 GTGATCTGGGGCAAGCCCCCAAGTTCGCGCTCCCATCCAGAAGAGAGACTCGGAGACC 1872
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
960 ValLeuTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979
QY 1873 TGTGACACGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1932
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProLysTrpGluPheValAsnThrPro 999
QY 1933 CCCTCTGTGAAGCTGTGTACCAAGCTGGAGAAGGAGCCCATCATCGCGCGCGAGACCTTC 1992
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019
QY 1993 TAGCTGGACGGCGCCCAACCGCAGAGACCAAGATCGGCAAGCGCGCTACGTGACCCGAC 2052
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyCysValThrAsp 1039
QY 2053 CGGGCGCGCAGAGAAGTCTGTGACCTGACCGAGACCAACACCAAGAGACCGGAGCTGCAG 2112
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059
QY 2113 GCCATCCAGCTGSCCTCGCAGGACAGCGCAGGTGAACATCGTGACCGACAGCCAG 2172
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079
QY 2173 TAGCCCTGGGCATCATCCAGGCCCGCCAGCAAGAGCGAGCGAGCTGGTGAACCCAG 2232
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099
QY 2233 ATCATCGACGAGCTGTATCAAGAGGAGAGGCTTACTGAGCTGGGTGGCGCCGCCACAAG 2292
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119
QY 2293 GGATCTCGCGGCAACGACGAGATCGCAAGCTCGTGAGCAAGGCGCATCCCAAGGTGCTG 2352
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1120 GlyIleGlyHisGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139
QY 2353 TTCCTGACGCGCATCGAT 2370
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1140 PheLeuAspGlyIleAsp 1145

RESULT 11
Q9WF65 PRELIMINARY; PRT; 1427 AA.
ID Q9WF65 AC Q9WF65;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.;
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana.";
RL J. Virol. 73:4427-4432 (1999).
```



1273 AGAGTCGAGGAGCTGGCAAGCACCTGCTCGCTGGGGCTTACACCCCGACAAAG 1332  
 Db |||||GluGluLeuArgGluHisLeuLeuLeuTyrGlyLeuThrThrProTyrLysLys 799  
 1333 CACCAGAGGAGCCCTCTCTGTGGATGGGTACGAGCTGACCCCGACAAGTGAC 1392  
 Db |||||GlnLysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyrThr 819  
 1393 GTGAGCCATCAGCTGCGCGAGAGAGAGAGCTGACCTGAACGACATCCAGAGCTG 1452  
 Db |||||GlnProIleGlnLeuProAspLysAspSerTyrThrValHisAspIleGlnLysLeu 839  
 1453 GTGGCAAGCTGAACCTGGGCGCCAGATCTACCCGGCATCAAGTGGCCAGCTGTGC 1512  
 Db |||||GlnLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuLys 859  
 1513 AAGCTGTGCGCGCGCAAGGCCCTGACGACATCGTCCCTGACCGAGAGGCGGAG 1572  
 Db |||||GlnLysLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu 879  
 1573 CTGAGCTGCGCGAGAACCGGAGATCTCTGGGAGCCCTGACCGCTGACTAGCAC 1632  
 Db |||||GlnLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp 899  
 1633 CCAGCAAGGACCTGTGGCGCGAGATCCACAAGCAGGCGCCACGACAGTGCACCTAC 1692  
 Db |||||GlnLysAspLeuIleAlaGluValGlnLysGlnIleHisAspGlnTyrThrGln 919  
 1693 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCACCGCC 1752  
 Db |||||GlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
 1753 CACACCAAGCTGAAGAGCTGACCGAGCGCTGCAGAAAGATCGCCATCGAGAGCATC 1812  
 Db |||||GlnAsnAspValLysGlnLeuThrGluValGlnLysIleThrMetGluSerIle 959  
 1813 GTGATCTGGGCGAAGCCCAAGTCTCGCTGCCATCCAGAAAGAGACCTGGGAGACC 1872  
 Db |||||GlnLysThrProLysPheArgLeuProIleGlnLysAspThrTyrGluThr 979  
 1873 TGGTGAACGAGTACTGCGAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACCCC 1932  
 Db |||||GlnThrAspTyrTyrGlnAlaThrTyrIleProGluTyrGluPheValAsnThrPro 999  
 1933 CCGCTGTGAAGCTGGTACGAGTGGAGAGAGCCCATCATCGCGCGCGAGACCTTC 1992  
 Db |||||GlnValLysLeuTyrTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019  
 1993 TAGTGGACGCGCGCGCAAGCGGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2052  
 Db |||||GlnAspGlyAlaAlaAsnArgGluThrLysLysAlaGlyTyrValThrAsp 1039  
 2053 CGGCGCGCGCAGAGATCTGAGCCTGACCGAGACCAACAGAGAGACCGAGCTGCAG 2112  
 Db |||||GlnArgGlnLysIleValProLeuThrGluThrThrAsnGlnArgAlaGluLeuGln 1059  
 2113 GCATTCAGCTGGCTGCGAGGAGCGGAGGAGTGAACATCTGTCGACCGAGCCAG 2172  
 Db |||||GlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
 2173 TAGCCTGCGGCATCATCCAGGCGCGCGCAAGAGCGAGCGAGCTGGTGAACAC 2232  
 Db |||||GlnLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099  
 2233 ATCATCAGAGCTGTGATCAAGAGGAGAGTGTACCTGAGCTGGTGGCGCGCCACAAG 2292  
 Db |||||GlnGluGlnLeuIleLysLysGluArgValTyrLeuSerTyrProAlaHisLys 1119  
 2293 GGCATCGCGCGCAAGAGAGAGATCGACAGCTGTGAGCAGGCGCATCCGCGAGGTCTG 2352  
 Db |||||GlyIleGlyAsnGluGlnValAspLysLeuValSerGlyIleArgLysValLeu 1139

2353 TTCTGCGCGCATCGAT 2370  
 Db |||||PheLeuAspGlyIleAsp 1145  
 RESULT 12  
 Q9WF77 PRELIMINARY; PRT; 1437 AA.  
 AC Q9WF77  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana.";  
 RL J. Virol. 73:4427-4432(1999).  
 CC -1- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AF110967; AAD17072.1; --  
 DR HSSP; P24736; INCP.  
 DR GO; GO:0019012; C:virion; IEA.  
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008907; F:integrase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0015074; P:DNA integration; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn\_N.  
 DR InterPro; IPR001995; Peptidase\_A2.  
 DR InterPro; IPR009007; Peptidase.  
 DR InterPro; IPR001969; Peptidase.  
 DR InterPro; IPR010999; Retroviral\_matrix.  
 DR InterPro; IPR000071; Retroviral\_p17.  
 DR InterPro; IPR008916; Retroviral\_capsid\_C.  
 DR InterPro; IPR008919; Retroviral\_capsid\_N.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF06815; RVT\_connect; 1.  
 DR Pfam; PF06817; RVT\_thumb; 1.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PRINTS; PR00234; HIVMATRIX.  
 DR SMART; SM00343; Znf\_C2HC; 2.

DR PROSITE; PS00141; ASP PROTEASE; 1.  
DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS0158; ZF\_CCHC; 2.  
KW AID5; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
KW Zinc-finger.  
SQ SEQUENCE 1437 AA; 161978 MW; 65DAA2562FC317A6 CRC64;



Db	480	PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlycIle	499
Qy	433	AAGGAGGCGCTGCTGGACACACGGCGCCGACACACCGTGTGGAGGAGATGAGCTGCC	492
Db	500	ArgGluAlaLeuLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro	519
Qy	493	GGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGAGTAC	552
Db	520	GlyLysTrpLysProLysMetIleGlyGlyPheIleLysValArgGlnTrp	539
Qy	553	GACCAGATCTTCATCGAGATCTCGGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCC	612
Db	540	GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro	559
Qy	613	ACCCCGGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCTCAACTTC	672
Db	560	ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe	579
Qy	673	CCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCGGATGAGCGGCCCAAG	733
Db	580	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	599
Qy	733	GTGAAGCAGTGCCTCCACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGCAGGAG	792
Db	600	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp	619
Qy	793	ATGAGAAAGAGGGCAAGATCACAAAGATCGGCGCCGAGAACCCCTACAACACCCCGTG	852
Db	620	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal	639
Qy	853	TTCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGCAGCTG	912
Db	640	PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
Qy	913	AACAAGCCACCCAGGACTTCTGGAGGTGTCAGCTGGGCATCCCGCCACCCCGCGCGCTG	972
Db	660	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	679
Qy	973	AAGAAGAAAGAGCGTCACCGTGTGGACGTGGCGCAGCGCTACTTTCAGCGTCCCGCTG	1032
Db	680	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
Qy	1033	GACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAAGAGACCCCC	1092
Db	700	AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro	719
Qy	1093	GGCATCCGCTACCAAGTCAACAGTCTGCCCGAGCGCTGGAGGGCAGCCCGCAGCATCTTC	1152
Db	720	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	739
Qy	1153	CAGACGACATGACCAAGATCTCGAGCGCTTCGCGCGCCGCAACCCGAGATCGTGATC	1212
Db	740	GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle	759
Qy	1213	TACCAGTACATGAGCACTCTACGTGGGCAAGCACTTGGAGATCGCGCCAGCACCGCGCC	1272
Db	760	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
Qy	1273	AAGATCGAGGAGCTGCGCAACCACTGCTGGCGCTGGGCTTCACACCCCGCAGCAAGAAG	1332
Db	780	LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys	799
Qy	1333	CACCAGAAGGAGCCCCCTCTGTGGATGGGCTACGAGCTGACCGCCGACCAAGTGGACC	1392
Db	800	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	819
Qy	1393	GTGCAGGCCCATCGAGCTGCCGAGAAGAGGAGCTGGACCGCTGAACGACATCCAGAAGCTG	1452
Db	820	ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu	839
Qy	1453	GTGGGCAAGCTGAAGTGGGCGAGCCAGATCTACCCCGGCAAGTCAAGTGGCGCAGCTGTC	1512
Db	840	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValHisLeuLysCys	859

QY	1513	AAGTGTGTCGGCGCGCAAGGGCCCTGACCGACATCTGTCGCCCTGACCGGAGGCGCCGAG	1572
Db	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu	879
QY	1573	CTGAGACTGCGCGAAGACCGCGAGATCTCTCGGCGAGCCCTGTGACGCGGTGTACTAGCAC	1632
Db	880	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	899
QY	1633	CCACGACGACCTGTGTGGCGGAGATCCAGAAGCAGGGCCACACGACGAGTGGACCTACCAG	1692
Db	900	ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln	919
QY	1693	ATCTACGAGGACCCCTTCAAGAACCTGGAAGACCGGCAAGTAGTACGCCAAGATGCGCACCGCC	1752
Db	920	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
QY	1753	CACACCAACGAGCTGAAGCAGCTGACCGAGCCGCTGCAGAGATCGCCATGGAGAGCAATC	1812
Db	940	HisThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle	959
QY	1813	GTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAGAGAGACCTCGGAGACC	1872
Db	960	ValIleTrpGlyLysThrProLysPheArgLysProLleGlnLysAspThrTrpGluThr	979
QY	1873	TGTGTGACCGACTACTGCGAGGGCCACCTGTGATCCCGAGTGGGAGTTCGTGAACACCCCC	1932
Db	980	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	999
QY	1933	CCCCTGTGAAGCTGTGGTACCAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTC	1992
Db	1000	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe	1019
QY	1993	TAGCTGACCGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGCTACGTGACCGAC	2052
Db	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	1039
QY	2053	CGGGCGCGGCAAGATCTGTGAGCGCTGACCGAGACCAACCAACAGAAAGACCGAGCTGCAG	2112
Db	1040	LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnLysAlaGluLeuGln	1059
QY	2113	GCAATCAGCTGGCCCTGCGAGGACAGCGGAGCGAGGTGAACATCTGTGACCGGACGCCAG	2172
Db	1060	AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln	1079
QY	2173	TAGCCCTGTGGCATCATCCAGGCCCGCCGACGAGCGAGCGAGCGAGCTGGTGAACACAG	2232
Db	1080	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln	1099
QY	2233	ATCATCAGCAGCTGTATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGCGCGCCCAACAG	2292
Db	1100	IleIleGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProIleHisLys	1119
QY	2293	GGCATCGCGCGCAACGAGCGATCGACAGCTGGTGAGCAAGGGCATCCGCAAGGTGTGTG	2352
Db	1120	GlyIleGlyAsnGlnGluValAspLysLeuValSerSerGlyIleArgLysValLeu	1139
QY	2353	TTCTCGGACGGCATCGAT	2370
Db	1140	PheLeuAspGlyIleAsp	1145
RESULT 14			
Q8ADX5			
ID	Q8ADX5	PRELIMINARY;	PRT; 1427 AA.
AC	Q8ADX5;		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Gag-pol fusion polypeptide (Fragment).		
GN	Name=gag-pol;		
OS	Human immunodeficiency virus 1.		
OC	Viruses; Retroviridae; Retroviridae; Lentivirus.		
OX	NCBI TaxID=11676;		





793 ATGAGAGGAGGCAAGATCACCAAGATCGGCGCCGAGAACCCCTACACACACCCCGTG 852  
Db MetGluLysGluGlyLysIleThrLysIleGlyProGluAenProTyrAsnThrProIle 639  
853 TTCCGCATCAAGAGAGGAGCAGCACAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912  
Db PheAlaIleLysLysLysAspSerThrLysTirpArgLysLysValAspPheArgGluLeu 659  
913 AACAGCGCACCCAGGACTTCTGGAGGTGAGTGGGCATPCCGCCACCCCGCGGCTG 972  
Db AsnLysArgThrGlnAspPheTirpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
973 AAGAAGAAGAGCGGTGACCGTCTCGAGCTGGAGCTGGGCACGCTACTTACGCGTCCCGTG 1032  
Db LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
1033 GACGAGACTTCCCAAGTACACCGCTTACCATCCCGCATCAACAGCATCAACAGGACACCCC 1092  
Db AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
1093 GGCAATCCGCTACCAAGTACAAGTCTGCTGCCAGGGCTGGAAAGGCGAGCCCGCATCTTC 1152  
Db GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTirpLysGlySerProAlaIlePhe 739  
1153 CAGAGCAGCATGACCAAGATCCTCGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759  
1213 TACCAGTACATGAGCAGCTCTAGTGGGCGAGCAGCTGGAGATCGCGCAGCACCGCGCC 1272  
Db TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779  
1273 AAGATCAGGAGCTGCGCAGCACCTGCTGGCTGGCTGGGGCTTACCACCCCGCAAGAAG 1332  
Db LysIleGluGluLeuArgGluHisLeuLeuArgTirpGlyPheThrThrProAspLysLys 799  
1333 CACAGAGAGGAGCCCGCTTCTCTGTGATGGGTACGAGCTGGACCCCGACCAAGTGAC 1392  
Db HisGlnLysGluProPhePheLeuTirpMetGlyTyrGluLeuHisProAspLysTirpThr 819  
1393 GTGACGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGATCCACCAAGCTG 1452  
Db ValGlnProIleGlnLeuProGluLysAspSerTirpValAsnAspIleGlnLysLeu 839  
1453 GTGGCGCAAGCTGAATCGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTC 1512  
Db 840 ValGlyLysLeuAenTirpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 859  
1513 AAGCTGTGCGCGCGCAAGGCCCTGACCGACATCGTGGCCCTGACCGAGGAGCGCGAG 1572  
Db LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluLysAlaGlu 879  
1573 CTGAGCTGGCGAGACCGGAGATCTCGCGCAGCCCGTGCAGCGCGTGTACTACGAC 1632  
Db LeuGluLeuAlaGluAenArgGluLeuLysGluProValHisGlyValTyrTyrAsp 899  
1633 CCCAGCAGGACCTGGTGGCGCAGATCCAGAAGCAGGCGCCACGACCATGAGCTACACAG 1692  
Db ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTirpTyrGln 919  
1693 ATTCACGAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1752  
Db IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939  
1753 CACACCAACGCTGAGCAGCTACCGAGCGCTGCAGAGATCGCATGGAGAGATC 1812  
Db HisThrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959  
1813 GTGATCTGGGCAAGACCCCAAGTTCGCTGCGCTGCCATCCAGAAGGAGACCTGGAGAC 1872  
Db IleIleTirpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTirpGluAla 979  
1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCC 1932

980 TrpTirpThrAspTyrTirpGlnAlaThrTirpIleProGluTirpGluPheValAsnThrPro 999  
1933 CCCCTGTGAAGCTGTGTACCTACGCTGGAGAAGAGCCCATCATCGCGCGCAGACCTTC 1992  
Db ProLeuValLysLeuTirpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019  
1993 TACGTGACGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2052  
Db TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039  
2053 CGGGCGCGCAGAGATCGTGCAGCTCACGAGACCAACAGAGAGAGCGAGCTGCAG 2112  
Db ArgGlyArgGlnLysIleIleSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
2113 GCATCCAGCTGGCGCTTCAGGAGCAGCGGAGGAGGAGTGAACATCGTGCAGCGACCGAC 2172  
Db AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValThrAspSerGln 1079  
2173 TACGCTCGGCGCATCTCCAGGCGCCGAGCCGAGCAAGAGCGAGCGAGCTGGTGAACGAC 2232  
Db TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099  
2233 ATCATCCAGCAGCTGATCAAGAGGAGAGTGTACCTGAGTGGTGGCGCGCCGCGACGCTG 2292  
Db IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTirpValProAlaHisLys 1119  
2293 GGATCCGCGCGCACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGCATCCGCAAGTGTCTG 2352  
Db GlyIleGlyLysGlnGluValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
2353 TTCTCGACGCGCATCGAT 2370  
Qy PheLeuAspGlyIleAsp 1145  
Db  
RESULT 15  
Q9WF79 PRELIMINARY; PRT; 1425 AA.  
ID Q9WF79 AC  
AC Q9WF79; (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-poi polypeptide.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RL J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110969; AAD17087.1; -.  
DR HSP; P24736; INCP.  
DR GO; GO:0019012; C:viroin; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003198; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0015074; P:DNA integration; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.



Db	838	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys	857
Qy	1513	AAGTGTCTGCGCGCCGAGCCCTGACACACATCGTCCCTGACCCAGGAGCCGAG	1572
Db	858	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProThrGluGluAlaGlu	877
Qy	1573	CTGAGCTGCGCGAGAACCGGAGATCTCGCGAGCCCGTGCACGGCGTGTACTACGAC	1632
Db	878	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp	897
Qy	1633	CCCAGCAGGACCTGCTGGCGAGATCCAGAGCAGGGCCACGACCTGACCTACCAG	1692
Db	898	ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspGlnTrpThrTyrGln	917
Qy	1693	ATCTACCAGGAGCCCTTCAAGACCTGAAGACCCGCGCAAGTACGCCAAGATCGCACCGCC	1752
Db	918	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	937
Qy	1753	CACACCAACGACGTGACGACCTGACCGAGCCCGTGCAGAGATCGCCATCGAGAGCATC	1812
Db	938	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle	957
Qy	1813	GTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAGAGACCTGGGAGACC	1872
Db	958	ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla	977
Qy	1873	TGCTGGACCCGACTACTGGCAGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCC	1932
Db	978	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	997
Qy	1933	CCCTGTGTGAAGCTGTGTACCACTGGAGAGAGCCCATCATCGGCGCCGAGACCTTC	1992
Db	998	ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe	1017
Qy	1993	TACGTGACCGCGCGCCCAAGCGGAGACCAAGATCCGCAAGCGCGGTACGTACCGAC	2052
Db	1018	TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp	1037
Qy	2053	CGGGCGCGGCAAGATCGTGACCTGACCCAGACCCACCAACAGAGACCCGAGCTGCAG	2112
Db	1038	LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln	1057
Qy	2113	GCCATCCAGCTGGCCCTGACGAGCAGCGGCGAGCGAGTGAAACATCGTGACCGAGCCAG	2172
Db	1058	AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln	1077
Qy	2173	TACGCCCTGGGCATCATCCAGCCCGCCGAGCCGAGAGCGAGAGCGAGCTGGTGAACCG	2232
Db	1078	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln	1097
Qy	2233	ATCATCGAGCAGCTGATCAAGAAGAGAGGTGTACCTGAGCTGGTGGCCGCCCAACAG	2292
Db	1098	IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys	1117
Qy	2293	GGCATCGCGCGCAACGACGACATCGAAGCTGGTGAAGAGGCGATCCCGCAAGGTGCTG	2352
Db	1118	GlyIleGlyLysAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu	1137
Qy	2353	TTCCTGGACGGCATCGAT	2370
Db	1138	PheLeuAspGlyIleAsp	1143

Search completed: June 2, 2005, 05:02:46  
Job time : 301.175 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 02:53:39 ; Search time 40.9328 Seconds  
(without alignments)

11607.276 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 4583

Sequence: 1 gtcagccaccatggccga.....gggctagcaccgtgaattc 2469

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2\_1/USFTO.spool/US09610313/runat\_31052005\_155137\_15159/app\_query.fasta\_1.7893  
-DB=PIR\_79 -QMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCHI=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CGN 1.189 @runat\_31052005\_155137\_15159 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR 79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3543.5	77.3	1002	2	S54378 pol polyprotein -
2	3535.5	77.1	1002	1	GNLJND HIV-1 retropepsin
3	3530	77.0	1003	1	GNVWJ3 HIV-1 retropepsin
4	3513	76.7	1015	1	GNVWJ3 HIV-1 retropepsin
5	3491	76.2	1003	2	T09440 pol polyprotein -
6	3488	76.1	1003	1	B44001 HIV-1 retropepsin
7	3487	76.1	1003	1	GNVWJ2 HIV-1 retropepsin
8	3477	75.9	1012	1	GNVWJ1 HIV-1 retropepsin
9	3249	70.9	1027	1	GNLJST HIV-1 retropepsin
10	3138	68.5	912	2	S33980 pol polyprotein -
11	3132	68.3	902	2	T01668 pol polyprotein -
12	2809	61.3	559	2	B47175 reverse transcript
13	2789	60.9	559	2	A47175 reverse transcript
14	2388	52.1	1039	2	S46347 pol polyprotein -

15	2244	49.0	1055	2	S53092 pol polyprotein -
16	2240.5	48.9	1055	1	GNLJST HIV-1 retropepsin
17	2231.5	48.7	1056	1	GNLJG3 HIV-1 retropepsin
18	2231	48.7	1036	1	GNLJG2 HIV-1 retropepsin
19	2223	48.5	1034	1	GNLJCA HIV-1 retropepsin
20	2222.5	48.5	1035	1	GNLJGG HIV-1 retropepsin
21	2212.5	48.3	1019	2	T11560 pol polyprotein -
22	2206.5	48.1	1009	2	S28081 pol polyprotein -
23	2185.5	47.7	1032	2	S12153 pol polyprotein -
24	2108	46.0	1058	2	S08436 pol polyprotein -
25	2088	45.6	1061	1	GNLJG4 HIV-1 retropepsin
26	2054.5	44.8	1054	1	GNLJG5 HIV-1 retropepsin
27	1998.5	43.6	656	2	S30484 pol polyprotein -
28	1969.5	43.0	656	2	S30483 pol polyprotein -
29	1428	31.2	1124	2	S23820 pol polyprotein -
30	1425	31.1	1124	1	GNLJFP HIV-1 retropepsin
31	1382	30.2	1124	2	B45557 HIV-1 retropepsin
32	1172	25.6	1146	1	GNLJ22 HIV-1 retropepsin
33	1172	25.6	1146	1	GNLJEW HIV-1 retropepsin
34	1169.5	25.5	1109	1	B45345 HIV-1 retropepsin
35	1152.5	25.1	1087	2	QJ1162 Pol protein - Maed
36	1152.5	25.1	1145	1	GNLJEB HIV-1 retropepsin
37	1149.5	25.1	1086	1	B46335 HIV-1 retropepsin
38	1142	24.9	1101	1	B45390 HIV-1 retropepsin
39	1138	24.8	1101	1	GNLJVS HIV-1 retropepsin
40	1128	24.6	216	2	S43127 reverse transcript
41	1109	24.2	219	2	S32152 RNA-directed DNA p
42	1105	24.1	219	2	S32118 RNA-directed DNA p
43	1105	24.1	219	2	S32139 RNA-directed DNA p
44	1103	24.1	219	2	S32089 RNA-directed DNA p
45	1103	24.1	219	2	S32160 RNA-directed DNA p

#### ALIGNMENTS

##### RESULT 1

S54378 pol polyprotein - human immunodeficiency virus type 1 (fragment)

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004

C;Accession: S54378

R;Theodore, T.; Buckler-White, A.J.

submitted to the EMBL Data Library, July 1989

A;Reference number: S54377

A;Accession: S54378

A;Status: preliminary

A;Molecule type: genomic RNA

A;Residues: 1-1002 <THE>

A;Cross-references: UNIPROT:P12499; EMBL:M22639; NID:G329377; PIDN:AAA5366.1; PID:G3293

C;Superfamily: pol polyprotein

C;Keywords: polyprotein

Alignment Scores:			
Pred. No.:	7.65e-156	Length:	1002
Score:	3543.50	Matches:	659
Percent Similarity:	96.25%	Conservative:	34
Best Local Similarity:	91.53%	Mismatches:	24
Query Match:	77.32%	Indels:	3
DB:	2	Gaps:	2

US-09-610-313B-30 (1-2469) x S54378 (1-1002)

QY	220	TTCTTCGCGGAGGACCTGGCGCTTCCCGAGGCGAAGCCCGCGATTCCCGAGGAGCAG 279
DB	1	PhePheargGluaspLeuAlaPheProGlnGlylsAlaGlyGluLeuSerSerGluGln 20
QY	280	AACGCGCCCAACAGCCCAACAGCCGAGCTGAGTGGCGGC---GACAAACCCCGC 336
DB	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40
QY	337	AGCGAGCCCGCGCCGAGCCGAGCCAGCCACCTGAACTTC-----CCCCAGTACCCCTG 390
DB	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu 60



```
QY 391 TGGCAGCGCCCTGTGTGACATCAAGGTGGCGGGCCAGATCAAGGAGGCCCTGTGGAC 450
Db 61 TrpGlnArgProLeuValThrIleLeuValIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80
QY 451 ACCGGCGCCGACACACACCGTGTGGAGGAGATGAGCCTGCCCGCAAGTGAAGCCCAAG 510
Db 81 ThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100
QY 511 ATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGGCAGTACGACACGATCCTGATCGAG 570
Db 101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu 120
QY 571 ATCTGGCGCAAGAGGCCATCGCAGCGTGTGATCGGCCCGCCACCCCGTGAACATCATC 630
Db 121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140
QY 631 GCGCGCAACATGTGACCCAGCTGGGTGACCCCTGAACCTTCCCATCGACCCCATCGAG 690
Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160
QY 691 ACCGTGCGCTGAAGTGAAGCCCGCATGGACGGCCCAAGGTGAAGCATGTGGCCCTG 750
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180
QY 751 ACCGAGGAGAGATCAAGGCCCTGACCCCATCTGCGAGGAGATGGAGAGGAGGGCAAG 810
Db 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200
QY 811 ATCACCAGATCGCGCCCGAGAACCCCTACAACACCCCGTGTGGCCATCAAGAGAAG 870
Db 201 IleSerArgValGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220
QY 871 GACAGACCAAGTGGCGCAAGCTGTGGACTTCCGGCGAGCTGAACAGCCGACCCAGGAC 930
Db 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240
QY 931 TTCTGGGAGGTGAGCTGGCGCATCCCGCCACCGCGCGCTGAAGAGAGAGAGAGCGTG 990
Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260
QY 991 ACCGTGCTGACGTGGCGCAGCGCTACTTTCAGCGTGCCTTGGACGAGGACTTCCGCAAG 1050
Db 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLys 280
QY 1051 TAGACCGCTTACCATTCCCGACGATCAACAGAGACCCCGCGCATCCGCTACCAGTAC 1110
Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300
QY 1111 AAGCTGCTGCCCGAGGCTGMAAGGCGAGCCCGACGATCTTCCAGAGCAGCATGACCAAG 1170
Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320
QY 1171 ATCTGGAGCGCTTCCCGCGCCGCAACCCCGAGATCGTGTATCTACCAAGTACATGAGCAG 1230
Db 321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340
QY 1231 CTGTACTGGCGCAGCACTTGGAGATCGCGCAGCACCAGCGCCCAAGATCGAGGAGCTCGC 1290
Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360
QY 1291 AAGCACTGTGCGTGGCGCTTCCACCACCCCGACAAAGACGACCAAGAGGAGCGCCCGC 1350
Db 361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380
QY 1351 TTCTGTGGATGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTG 1410
Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400
QY 1411 CCGAGAGAGAGAGCTGGACGCTGAACGACATCCAGAAAGCTGTGGCGCAAGCTGAACTGG 1470
Db 401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420
```

```
QY 1471 GCCAGCCAGATCTACCCCGCATCAAGGTGGCGGCAGCTGTCAAGCTGTGCGCGCGGCC 1530
Db 421 AlaSerGlnIleTyrProGlyLysValArgGlnLeuCysLysLeuLeuAspGlyThr 440
QY 1531 AAGGCTTGAACGACATCGTGGCCCTGACCGAGGAGGCCGAGCTGTGAGCTGGCGGCAAC 1590
Db 441 LysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGluAsn 460
QY 1591 CGCGAGATCTCGCGCAGCGCGTGCACGGGTGTACTACGACCCAGCAAGGACCTGGTG 1650
Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProserLysAspLeuIle 480
QY 1651 GCCGAGATCCAGAAAGAGGCGCCACGACCTGTGACCTTACCCAGATCTTACCAAGGACCCCTTC 1710
Db 481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500
QY 1711 AAGAACTGAAGACCGGCAAGTACGCAAGATCGGCAAGATCGGCAAGCCGACCAACAGCGTGAAG 1770
Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520
QY 1771 CAGCTGACCGAGCGCTGCAAGATCGCATCGAGAGCATCGTGTCTGGGCAAGACCC 1830
Db 521 GlnLeuAlaGluValValGlnLysIleSerThrGluSerIleValIleTrpGlyLysThr 540
QY 1831 CCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGTGTGACCGACTACTGG 1890
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpValGluTyrTrp 560
QY 1891 CAGGCAACTGTGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGG 1950
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580
QY 1951 TACGAGCTGAGAGAGAGCGCCATCATCGCGCGGAGACCTTCTACGTGGACCGCGCGGCC 2010
Db 581 TyrGlnLeuGlnLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600
QY 2011 AACCGGAGACCAAGATCGGAGCGCGCTACGTGACCGACCGCGCGCGCGCAAGATC 2070
Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620
QY 2071 GTGAGCTGACCGAGACCCCAACCGAGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130
Db 621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
QY 2131 CAGCAGACGCGCAGCAGGTGAACATCGTCACGACGACGAGTACGCGCTGGCGCATCATC 2190
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660
QY 2191 CAGCGCCAGCCGACAGAGCGAGCGAGCTGGTGAACGACATCATCGAGCAGCTCATC 2250
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680
QY 2251 AAGAGAGAAAGTGTACTGTGAGTGGTGGTCCCGCCGACAAAGGCGCATCGCGGCAAGCAG 2310
Db 681 LysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700
QY 2311 CAGATCGACAAGCTGTGTGACCAAGGCGATCCGGAAGGTGTGTCTTGGACGCGCATCGAT 2370
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
RESULT 2
GNLJND
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: JQ0067
R;Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.;
Gene 81, 275-284, 1989
A;Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunoc
A;Reference number: JQ0065; MUID:90034200; PMID:2806917
A;Accession: JQ0067
```

A:Molecule type: DNA  
 A:Residues: 1-1002 <SPI>  
 A:Cross-references: UNIPROT:P18802; GB:M27323; NID:G328154; PID:AAA44869.1; PID:G328158  
 C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
 C:Genetics:  
 A:Gene: pol  
 C:Superfamily: pol polyprotein  
 C:Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle  
 F:56-154/Product: retropepsin #status predicted <RTP>  
 F:80/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
 Pred. No.: 1,79e-155 Length: 1002  
 Score: 3535.50 Matches: 659  
 Percent Similarity: 96.25% Conservative: 34  
 Best Local Similarity: 91.53% Mismatches: 24  
 Query Match: 77.14% Indels: 3  
 DB: 1 Gaps: 2

US-09-610-313B-30 (1-2469) x GNLJND (1-1002)

Qy	220	TTCTTCGGCAGGACCTGGCCCTTCCCGGCAAGCCCGCGAGTTCCCGAGCGAGCAG	279
Db	1	PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluPheSerSerGluGln	20
Qy	280	AACCGCCCAACACCCCAACAGCCGCGAGCTGCAGGTG---CGCGCGACAACCCCGCG	336
Db	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu	40
Qy	337	AGCGAGCCGCGCGAGCGCGAGCGCGACCCCTG-----AACTTCCCGAGATCACCTG	390
Db	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu	60
Qy	391	TGGCAGCGCCCTCGGTGAGCATCAAGTGGCGGCGGCGAGTCAAGAGGCGCCCTCTCGAC	450
Db	61	TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp	80
Qy	451	ACCGGCCCAACGACACCGTCTCGAGAGATGAGCTGCCCGCAAGTGGAAAGCCCAAG	510
Db	81	ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys	100
Qy	511	ATGATCGCGCGCATCGCGGCTTTCATCAAGTGGCGGCGAGTACGACCGAGTCTGATCGAG	570
Db	101	MetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu	120
Qy	571	ATCTGCGCAAGAGGCATCGGCACCGTGTGATCGGCCCGCCACCCCGTGAACATCATC	630
Db	121	IleCysGlyTyrllysAlaMetGlyThrValLeuValGlyProThrProValAsnIleIle	140
Qy	631	GGCGCAACATGTGACCCAGCTGGGCTGCACCTGAACTTCCCATCAGCCCGCATCGAG	690
Db	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
Qy	691	ACCGTCCCGTGAAGTGAAGCCCGGATGACCGGCCCGCCCAAGTGAAGCATGTGCCCTG	750
Db	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
Qy	751	ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTCGCGAGGATGGAGAGGGGCAAG	810
Db	181	ThrGluGlnLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys	200
Qy	811	ATCACCAAGATCGGCCCGCGAGAACCCCTTACAACACCCCGCTGTTCGCCCATCAAGAAAG	870
Db	201	IleSerArgIleGlyProGluAsnProTyAsnThrProIlePheAlaIleLysLysLys	220
Qy	871	GACAGCAACAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAGCGGACCCAGGAC	930
Db	221	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgTrpGlnAsp	240
Qy	931	TTCTGGGAGTGGAGCTGGGCATCCCCACCCCGCGCTGAAGAGAGAGAGAGCGTG	990
Db	241	PhetrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	260

Qy	991	ACCGTGTGACGCTGGGCGAGCCCTACTTACGCTGCCCTGGACGAGGACTTCCGCAAG	1050
Db	261	ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluAspPheArgLys	280
Qy	1051	TACACCCGCTTACCATCCCGCAGCATCAACAACAGACCCCGCGCATCCGCTACCAGTAC	1110
Db	281	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrlGlnTy	300
Qy	1111	AACTGTGTGCCCGCAGGCTGGAAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAG	1170
Db	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
Qy	1171	ATCCTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGATCTACCATGATGACACAC	1230
Db	321	IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrlGlnTyrlMetAspAsp	340
Qy	1231	CTGTACGTGGCGCAGCAGCTGGAGATCGGCAGACACCGCGCAAGATCGAGGAGCTGGCG	1290
Db	341	LeuTyrlValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	360
Qy	1291	AAGCACCTGTCTGGCTGGGCTTCCACCAACCCCGCAGAACAGAACAGAGAGAGCCGCC	1350
Db	361	GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	380
Qy	1351	TTCTGTGTGATGGGCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTG	1410
Db	381	PhelLeuTrpMetGlyTyrlGluLeuHisProAspLysTrpThrValGlnProIleAsnLeu	400
Qy	1411	CCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGCAAGCTGCTCGCGGCGCC	1470
Db	401	ProGlnLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	420
Qy	1471	GCACGACAGATCTACCCCGCGCATCAAGTGTGGCGAGCTGTGCAAGCTGCTCGCGGCGCC	1530
Db	421	AlaSerGlnIleTyrlAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr	440
Qy	1531	AAGCCCTGACCGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1590
Db	441	LysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn	460
Qy	1591	CGCGAGATCTGCGGAGCCGCTGACGGGTGTACTACGACCCCGCAGCAAGCAAGCTGGTG	1650
Db	461	ArgGluIleLeuLysGluProValHisGlyValTyrlAspProSerLysAspLeuIle	480
Qy	1651	GCGGAGATCCAGAGCAGGCGCCACGACGAGTGGACCTTACAGATCTACGAGAGCCCTTC	1710
Db	481	AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrlGlnIleTyrlGlnGluProPhe	500
Qy	1711	AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCGCCACACCAACAGCGTGAAG	1770
Db	501	LysAsnLeuLysThrGlyLysTyrlAlaArgThrArgGlyAlaHisThrAsnAspValLys	520
Qy	1771	CAGTGAACGAGCGCTGCGAGAGATCGCCATGAGAGAGCATCTGATCTGGGCGCAAGACC	1830
Db	521	GlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr	540
Qy	1831	CCCAAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1890
Db	541	ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrlTrp	560
Qy	1891	CAGGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGGTGG	1950
Db	561	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp	580
Qy	1951	TACCATGTGAGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC	2010
Db	581	TyrGlnLeuGlnLysGluProIleGlyAlaGluThrPheTyrlValAspGlyAlaAla	600
Qy	2011	AACCGCAGAGCAAGATCGGACCGCGCTGCTGACCGCACCGCGCGCGCGCGAGCATC	2070
Db	601	AsnArgGluThrLysLeuGlyLysAlaGlyTyrlValThrAspArgGlyArgGlnLysVal	620
Qy	2071	GTGAGCCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCGCATTCAGTGGCCCTG	2130

```

Db      621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY      2131 CAGGACAGCGCGAGCGAGGTGAACATCGTACCCAGCAGCGAGTACGCGCTTGGGCATCATC 2190
Db      641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660
QY      2191 CAGGCCAGCCCAACAAGACGAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGAGCTGATC 2250
Db      661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680
QY      2251 AAGAAGGAGAGGTGTACCTGAGCTGGCTCCGCCCCACAAAGGCGCATCGCGCGCAACGAG 2310
Db      681 LysLysGlnLysValTyrLeuAlaIleValProAlaHisLysGlyIleGlyIleGlnGlu 700
QY      2311 CAGATCGACAAGCTGTGTGACAAAGGCGATCCGCAAGGTGTCTTCTGGACGCGCATCGAT 2370
Db      701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 3
GNVWL
N;Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 03-Jun-2002
C;Accession: A03966
R;Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A;Title: Nucleotide sequence of the AIDS virus, LAV.
A;Reference number: A90866; MUID:85099333; PMID:2981635
A;Accession: A03966
A;Molecule type: DNA
A;Residues: 1-1003 <WAI>
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
A;Gene: pol
C;Superfamily: pol polypeptide
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:
Pred. No.: 3,2e-155 Length: 1003
Score: 3530.00 Matches: 659
Percent Similarity: 95.98% Conservative: 33
Best Local Similarity: 91.40% Mismatches: 25
Query Match: 77.02% Indels: 4
DB: 1 Gaps: 2

US-09-610-313B-30 (1-2469) x GNVWL (1-1003)
QY      220 TTCTTCGCGAGGACCTGGCGCTTCCCGCAGGCAAGCGCGCGAGTTCCCGCAGCGAGCAG 279
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY      280 AACCGGCCCAACAGCCCAACAGCGCGAGCTGAGGTG-----CGCGCGCAACACCC 333
Db      21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTyrGlyArgAspAsnAsnSer 40
QY      334 CGCAGCGAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
Db      41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY      388 CTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
Db      61 LeuTyrGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY      448 GACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
Db      81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTyrLysPro 100
QY      508 AAGTGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567

```

```

Db      101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
QY      568 GAGATCTGGCGCAAGAGCCATCGGCGACCGTCTGATCGGCGCCACCCCGCTGGAACATC 627
Db      121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY      628 ATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACCTCCCATCAGCCCGCATC 687
Db      141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY      688 GACACCGTGGCTGAGCTGAAGCCCGCGCATCGAGCGCGCCCAAGGTGAAGCATGTGCC 747
Db      161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTyrPro 180
QY      748 CTGACCGAGAGAGATCAAGCCCTGACCGCCCATCTCTGAGGAGAGATGAGAGAGAGGCG 807
Db      181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY      808 AGATCACCAAGATCGCGCGCGCGAGAACCCCTACAACACCCCGCTTTCGCGCATCAAGAAG 867
Db      201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY      868 AAGCAGCAGCACCAAGTGGCGCAAGCTGTGTGGAATTCGCGAGCTGAACAAGCGCACCCAG 927
Db      221 LysAspSerThrLysTyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY      928 GACTTCTGGAGGTGAGCTGGCGATCCCGCCACCGCGCGCGCTGAGAGAGAGAGAGAGAG 987
Db      241 AspPheTyrGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY      988 GTGACCGTCTGAGCTGGCGGAGCGCTACTTCCAGGTGCGCGCTGAGAGAGAGAGAGAG 1047
Db      261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280
QY      1048 AAGTACACCGCTTCCACCATCCCGCAGCATCAACAAGCAGACCCCGCGCATCCGCTACAG 1107
Db      281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY      1108 TACAAGTGTGCTGCCAGGCTGGAAGGAGCGCGCGCGAGCATCTTCAGAGCAGCATGACC 1167
Db      301 TyrAsnValLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY      1168 AAGATCTGTGAGCCCTTCCGCGCGCGCGCAACCCCGAGATCGTGATCTACAGTACATGAC 1227
Db      321 LysIleLeuGlnProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY      1228 GACCTGTACGTGGCGCAGCACCTTGGAGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1287
Db      341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY      1288 CGCAAGACCTGTGCTGGCGCTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1347
Db      361 ArgGlnHisLeuLeuArgTyrGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380
QY      1348 CCTTCTCTGTGATGGCTACGAGCTGCACCCGACAGTGGACCGCGCGCGCGCGCGCGCGCG 1407
Db      381 ProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyrThrValGlnProIleVal 400
QY      1408 CTGCGCGCAGAGAGAGCTGACCGCTGAAACAGCATCCAGAAAGCTGTGGGCAAGCTGAAC 1467
Db      401 LeuProGluLysAspSerTyrThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY      1468 TGGCGCAGCGCAGATCTACCCCGCGCATCAAGGTGGCGCGCGCGCGCGCGCGCGCGCG 1527
Db      421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY      1528 GCCAAGCGCTGACCGGACATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1587
Db      441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuAlaGlu 460
QY      1588 AACCGGAGATCTCTGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1647

```



```
Db 281 AlatyPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
|||
QY 1072 AGCATCAACACGAGACCCCGCATCCGCTACAGTACAACTGCTGCCCGCAGGCTGG 1131
|||
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
|||
QY 1132 AAGGGACGCCCGCATCTTCCAGACGAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
|||
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
|||
QY 1192 CGAACCCCGAGATCGTGATCTACAGTACATGACGACTGTACGTGGCGACGACCTG 1251
|||
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
|||
QY 1252 GAGATCGCCGAGACCCCGCCAGATCGAGAGCTGGCGAAGCACCTGCTGCCCTGGGC 1311
|||
Db 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
|||
QY 1312 TTCACACCCCGCAGAACAGACACCAAGAGAGCCCTTCTGTGGATGGGTACGAG 1371
|||
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
|||
QY 1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGCTGAGCC 1431
|||
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
|||
QY 1432 GTGAACACATCCAGAGCTGTGGGCAAGCTGAATGGGCCAGCCAGATCTACCCCGGC 1491
|||
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
|||
QY 1492 ATCAAGTGTGCCAGCTGTGCAAGCTGCTCGCGCGCCCAAGCCCTACCGACATCGTG 1551
|||
Db 441 IleIysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
|||
QY 1552 CCCTGTACCGAGAGGCCGAGCTGGAGCTGCCGAGAACCGCGAGATCTCGCGAGGCC 1611
|||
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLysGluPro 480
|||
QY 1612 GTGCACGGCTGTACTACGACCCCGCAGCAAGACTGTGTGGCCGAGATCCAGAGACGGCC 1671
|||
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
|||
QY 1672 CACGACCATGGACTACACAGATCTACACAGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
|||
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
|||
QY 1732 TAGCCCAAGATGCGCACCCCGCCACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAG 1791
|||
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
|||
QY 1792 AGATCCCATGGAGACATCGTGATCTGGGGCAGACCCCAAGTTCGCGCTGCCCATC 1851
|||
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
|||
QY 1852 CAGAGGAGACCTGGGAGACTGTGTGACCGACTACTTGGCAGGCGCACTGATCCCGAG 1911
|||
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
|||
QY 1912 TGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACCACTGGAGAGAGGCC 1971
|||
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuLeuTrpTyrGlnLeuGluLysGluPro 600
|||
QY 1972 ATCATCGCGCGAGACTTCTAGCTGAGCGCGCCCGCCCAACCGCGAGACCAAGATCGGC 2031
|||
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
|||
QY 2032 AAGCCGGCTAGGTGACCGACCGGGCGCGCAGAAAGATCGTGAGCCCTGACCGACCAACC 2091
|||
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
|||
QY 2092 AACCAAGACCGAGCTGCAAGGCCATCCAGCTGGCCCTCGAGGACACCGCGAGCGAGGTG 2151
|||
```

```
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
|||
QY 2152 AACATCGTCACCCAGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGACCAAGAGC 2211
|||
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
|||
QY 2212 GAGACGAGCTGGTGAACAGATCATCGACGAGCTCATCAAGAGGAGAGGTGTACCTG 2271
|||
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
|||
QY 2272 AGCTGGTGGCCCGCCACAGGCGATCGCGGCAAGCAGCAGATCCACAGCTGTGAGC 2331
|||
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
|||
QY 2332 AAGGCGATCGCAAGCTGTCTCTCGACGCGATCGAT 2370
|||
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
|||
RESULT 5
T09440
pol polyprotein - human immunodeficiency virus type 1 (strain JRFL) (fragment)
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
R:Accession: T09440
R:Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A:Reference number: Z16673
A:Accession: T09440
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1003 <PAN>
A:Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779
C:Gene: pol
C:Superfamily: pol polyprotein
Alignment Scores:
Pred. No.: 1.99e-153 Length: 1003
Score: 3491.00 Matches: 650
Percent Similarity: 95.56% Conservative: 39
Best Local Similarity: 90.15% Mismatches: 28
Query Match: 76.17% Indels: 4
DB: 2 Gaps: 2
US-09-610-313B-30 (1-2469) x T09440 (1-1003)
QY 220 TTCTTTCGGCAGGACTGGCCCTTCCCGCAGGCAAGCCCGAGTTCCCGCAGGAGCAG 279
|||
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
|||
QY 280 AACCGCGCCAAACAGCCCGCCAGCGAGCTCGAGTG-----CGCGCGCAACACCC 333
|||
Db 21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTrpGlyArgAspSerAsnSer 40
|||
QY 334 CGCAGGAGCGCCCGCGCCGAGCGCCAGGGC-----ACCCTGAACCTTCCCGAGATCAC 387
|||
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyAsnValSerPheAsnPheProGlnIleThr 60
|||
QY 388 CTGTGGCAGGCCCGCTGTGACATCAAGTGGCGGCAGATCAGGAGGCGCCCTGCTG 447
|||
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
|||
QY 448 GACACCGCGCCCGCAGCACCGCTGTGGAGGAGATGAGCTGCGCGCAAGTGAAGCCC 507
|||
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
|||
QY 508 AAGATGATCGCGCGCATCGCGCTTCATCAAGGTGGCGCCAGTACGACCAAGTCTGATC 567
|||
Db 101 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
|||
QY 568 GAGATCTGGCGCAAGAGCGCATCGGCATCGCTGATCGGCCCGCCACCCCGTGAACATC 627
|||
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyThrValLeuValGlyProThrProValAsnIle 140
|||
```

```
QY 628 ATCGGCGCAACATGCTGACCCAGCTGGCTGCAACCTTGAACTTCCCATCAGCCCATC 687
Db 141 IleglyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCTGTCGCGTGAAGCTGAAGCCCGGATGACGCGCCCAAGGTGAAGCAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAAAGATCAAGGCGCTGACCGCCATCTCGAGGAGATGAGAAAGAGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCGCGGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCAACCAAGTGGCGCAAGCTGTGACCTTCCGAGCTGAACAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysLysThrGln 240
QY 928 GACTTCTGGGAGGTGACGTGGGCATCCCCACCCCGCGCTGAAGAAGAAAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysArgLysSer 260
QY 988 GTGACCTGTGACGTGGCGGCGACCTTACTTACAGCTGCCCTGGGACGAGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCTTACCATCCCGACGATCAACACAGAGACCCCGGCATCCGCTACCG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGlnThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGTGCCCCCAGGGCTGAAGGCGAGCCCGACGATCTTCCAGAGCAGCATGCC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCTCGAGCCCTTCCGCGCCCGCAACCCCGAGATGTGATCTACAGTACATGGAC 1227
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340
QY 1228 GACCTGTACCTGGCGGACCTGGAGATCGGCAGCAGCCGCGCAGAGTCCGAGGAGCTG 1287
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeu 360
QY 1288 CGCAAGCACCTGCTGCGCTGGGCTTCAACACCCCGCACAGAGACCAAGAGAGGCC 1347
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1348 CCCTTCTGTGGATGGCTACGAGCTGCACCCGACAAAGTGGACCGTGCAGCCCATCGAG 1407
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1408 CTGCCCCAGAGAGAGACTGGACCGTGAACGACATCCAGAGCTGGTGGCGAGCTGAAC 1467
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1468 TGGGCCAGCAGATCTACCCGCGCATCAAGTGGCGGCGAGCTGTGCAAGCTGCTCGCGGC 1527
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
QY 1528 GCCAAGCCCTGACCGACATGTTGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAG 1587
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1588 AACCGCAGATCTCGCCGAGCCCGTGCACGGGTGTACTACGACCCCGCAGGAGCTG 1647
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480
QY 1648 GTGCGCCAGATCCAGAGCAGGCGCACGACAGTGGACCTACCAAGTCTACCGAGGCC 1707
Db 481 IleAlaGluLeuGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
```

```
QY 1708 TTCAAGAACCTGAGACCGCGCAAGTACGCCAAGATGCGCACCGCCACCAACGACGTG 1767
Db 501 PheLysIleLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGGCGGTGCAAGATGCCATGCCATGAGAGCATCTGTGATCTGGGGCAAG 1827
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
QY 1828 ACCCCCAAGTTCGCGCTGCCATCCAGAAAGAGACCTGGGAGACCTGGTGACCGACTAC 1887
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
QY 1888 TGGCAGGCGCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGCTGAAGCTG 1947
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1948 TGGTACCAGCTGGAGAGAGGCCATCATCGGCGCCGAGACCTTCTACGTGGACGCGGCC 2007
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2008 GCCAACCGCAGACCAAGATCGCAAGCGCGCTTACGTGACCGACCGGGCGCGCAAG 2067
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2068 ATCGTGACCTGACCGAGACCAACCAAGACCGAGCTGCGAGGCACTCCACTGGCC 2127
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2128 CTGCAAGACCGCGCAGGAGTGAACATCGTACCGACAGCAGCAGTACGCGCTGGGATC 2187
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGCGCCAGCCGCAAGAGCAGAGCAGCTGTTGTAACCAAGATCATCGACGACTG 2247
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGAGAGAGTGTACTCTGAGCTGGTGGTGGCCGCCACCAAGGCACTCGGCGAAC 2307
Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700
QY 2308 GAGCAGATCCACAAGCTGGTGAGCAGGCGCATCGCAAGGTGCTGTCTTCTGGACGGCATC 2367
Db 701 GlnGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db 721 Asp 721

RESULT 6
B44001
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (strain YU-2)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: B44001
J;Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.
R;Virol. 66, 6587-6600, 1992
A;Title: Complete nucleotide sequence, genome organization, and biological properties of
A;Reference number: A44001; MUID:93021387; PMID:1404605
A;Accession: B44001
A;Molecule type: DNA
A;Residues: 1-1003 <11>
A;Cross-references: UNIPROT:P35963; GB:M93258
C;Comment: This protein is synthesized as a gag-pol polyprotein.
C;Comment: Specific enzymatic cleavages may yield mature proteins including proteinase,
C;Genetics:
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status predicted
```



Alignment Scores:

Pred. No.: 2,74e-153 Length: 1003  
 Score: 3488.00 Matches: 651  
 Percent Similarity: 95.42% Conservative: 37  
 Best Local Similarity: 90.29% Mismatches: 29  
 Query Match: 76.11% Indels: 4  
 DB: 1 Gaps: 2

US-09-610-313B-30 (1-2469) x B44001 (1-1003)

Qy	220	TTCTTCCGCGAGGACCTGGCTTCCCTCCCGAGGCAAGCCCGCGAGTTCCCGAGGAGCAG	279
Db	1	PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln	20
Qy	280	AACCGCGCAACAGCCACCACGCGCGAGCTCGAGTG-----CGCGCGCAACACCC	333
Db	21	ThrArgAlaAsnSerProIleArgGluGluArgGlnValTrpArgArgAspAsnAsnSer	40
Qy	334	CGCAGCGAGCGCGCGCGAGCGCCAGGCGCACCTG-----AACTTCCCGCGAGTCAACC	387
Db	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr	60
Qy	388	CTGTGGCAGCGCCCTGTGTGAGATCAAGTGGCGCGCGAGATCAAGAGGCGCTGTG	447
Db	61	LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu	80
Qy	448	GACACGCGCGCGAGCACCGTGTGGAGGAGATGAGCTGCGCGCAAGTGGAGGCC	507
Db	81	AspThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro	100
Qy	508	AAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGCGCAGTACGACACGATCCTGATC	567
Db	101	LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleProIle	120
Qy	568	GAGATCTGCGGCAAGAGGCATCGCACCGTGTGATCGGCGCCCGCGCGAGTGGCAATC	627
Db	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
Qy	628	ATCGCGCGACATGTGACCCAGCTGGCTGGCTGACCTGACCTGAACCTCCCGCATCGCCCATC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
Qy	688	GAGACCGTGCCTGGAAGCTGAAGCCCGCATGACCGCGCCCAAGGTGAAGCAGTGGCC	747
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
Qy	748	CTGACCGAGGAGAGATCAAGCGCCTGACCGCCATCTGCGAGGAGATGAGAGAGGGC	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
Qy	808	AAGATCACCAAGATCGGCGCGGAGAACCCCTACACACCCCGTGTTCGCGCATCAAGAAG	867
Db	201	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	220
Qy	868	AAGGACACCAAGTGGCGCAAGCTGGTGAGATTCGCGAGCTGAAACAAGCGCACCCAG	927
Db	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
Qy	928	GACTTCTGGAGGTGACGTGGGATCCCGCATCCCGCGCGCTGGAAGAGAGAGAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
Qy	988	GTGACCGTGTGGACGTGGCGGCGCTTCTGACGTGGCTGGACGAGGACTTCCGC	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuHisGluAspPheArg	280
Qy	1048	AAGTACACCGCTTCAACATCCCGCATCAACAACGAGACCCCGCGCATCCGCTACCA	1107
Db	281	LysThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyThrArgTyrGln	300
Qy	1108	TACACGTGTGCGCGAGGCTGAAGGCGAGCGCCAGCATCTTCCGACGAGCATGACC	1167
Db	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320

Qy	1168	AAGATCTGTGAGCCCTTCGCGCGCCCGCAACCCCGAGATCGTGTATCCAGTAGTACATG	1227
Db	321	ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTyrGlnTyrMetAsp	340
Qy	1228	GACCTGTACGTGGGCGAGCACTGTGAGATCGCGCAGCACCGCGCCCAAGATCGAGGAGCTG	1287
Db	341	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
Qy	1288	CGCAAGCACCTGTGTGGCTGGGGCTTCCACCCCGCAAGAGAGACACCAAGAGGAGCC	1347
Db	361	ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
Qy	1348	CCCTTCTGTGATGGCTTACGAGCTGCACCCCGCAAGTGGACCGCTGCGCCCATCGAG	1407
Db	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal	400
Qy	1408	CTCCCGAGAGAGAGAGCTGGACCGGTGAACGACATCAGAGCTGGTGGGCAAGCTGAAC	1467
Db	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
Qy	1468	TGGCGCAGCCAGATCTACCCCGCATCAAGGTGCGCGCAGCTGTCAAGCTGTCTGCGCGC	1527
Db	421	TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly	440
Qy	1528	GCCAAGCGCCCTGACCCGACATCTGTGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAG	1587
Db	441	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuAlaGlu	460
Qy	1588	AACCGGAGATCTTGGCGAGCCGTGCGCGGTGTACTACACCCCGAGAGGAGCTG	1647
Db	461	AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu	480
Qy	1648	GTGCGCGAGATCCAGAGGAGGCGCCACGACCTGAGCTGTACACGATCTACACGAGGACCC	1707
Db	481	IleAlaGluIleGlnLysGlnGlyGlnTrpThrTrpThrTrpThrTrpThrTrpThrTrpThr	500
Qy	1708	TTCAAGAACCTGAAAGACCGGCAAGTACGCGACCGCACCGCCCGCCACACACGAGCTG	1767
Db	501	PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal	520
Qy	1768	AAGCAGCTGACCGAGCGCTGTGAGAAGATCGCCATCGAGAGCATCTGTGATCTGGGGCAAG	1827
Db	521	LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys	540
Qy	1828	ACCCCAAGTTCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTAC	1887
Db	541	ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrTrpThrTrpThr	560
Qy	1888	TGCGAGCGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGTGGTGAAGCTG	1947
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu	580
Qy	1948	TGTTACAGCTGGAAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGGCC	2007
Db	581	TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla	600
Qy	2008	GCCAAACCGGAGACCAAGATCCGCAAGCGCGGTACGTGACCGACCGGGCGCGCAGAAG	2067
Db	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnLysGlyLysGlnLys	620
Qy	2068	ATCGTGTGAGCTGACCGAGCACCAACACCAAGAGACCGAGCTGCGAGGCGCATCCAGCTG	2127
Db	621	ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla	640
Qy	2128	CTGCGAGACAGCGCGAGGAGTGAACATCGTACCGACGAGCAGCTACGCGCTGGGATC	2187
Db	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle	660
Qy	2188	ATCCAGCGCCAGCCCGACAGAGGAGCGAGCTGTGTGAACCGACATCATCGAGCAGCTG	2247
Db	661	IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleIleGluGlnLeu	680

QY 2248 ATCAAGAGGAGNAGGTCTACTGAGCTGGGTGCCCGCCCAAGAGGCATCGCGGCAAC 2307  
Db 681 lilelyslsGluLysValTyrLeuAlaTrpValProAlaHisLysGlyGlyAsn 700  
QY 2308 GAGCAGATCGAAGCTGGTGAAGGCGATCCGCAAGGTGCTTCTCGACGGCATC 2367  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyTle 720  
QY 2368 GAT 2370  
Db 721 Asp 721  
RESULT 7  
GNVW2  
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate ARV-2)  
N;Contents: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
C;Species: human immunodeficiency virus type 1, HIV-1  
A;Note: host Homo sapiens (man)  
C;Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004  
C;Accession: A03968  
R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh  
Science 227, 484-492, 1985  
A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).  
A;Reference number: A04003; MUID:85090453; PMID:2578227  
A;Accession: A03968  
A;Molecule type: DNA  
A;Residues: 1-1003 <SAN>  
C;Cross-References: UNIPROT:P03369; GB:K02007; NID:g328658; PIDN:AAB59876.1; PID:g328662  
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
C;Genetics:  
A;Gene: pol  
C;Superfamily: pol polyprotein  
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle  
F;57-147/Product: retropepsin #status predicted <RTP>  
F;81/Active site: Asp (shared with dimeric partner) #status predicted  
Alignment Scores:  
Pred. No.: 3.04e-153 Length: 1003  
Score: 3487.00 Matches: 650  
Percent Similarity: 95.42% Conservative: 38  
Best Local Similarity: 90.15% Mismatches: 29  
Query Match: 76.09% Indels: 4  
DB: 1 Gaps: 2  
US-09-610-313B-30 (1-2469) x GNVW2 (1-1003)  
QY 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGCAAGCCCGCGAGTTCCCGAGGAGCAG 279  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 280 AACCGCGCCAAACGCCCCACCGAGCTGCGAGTGGCGCGGC-----GACAACCCC 333  
Db 21 ThrArgAlaAsnSerProThrArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40  
QY 334 CGAGAGGCGCGCGGCGAGCGCCAGGCGACCTG-----ACTTCCCGCAGATCACC 387  
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
QY 388 CTGTGGCAGCGCCCTGGTGAGCATCAAGGTGGCGCGCAGATCAAGAGGCGCCCTGCTG 447  
Db 61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80  
QY 448 GACACCGCGCGCAGACACCGCTGCTGGAGGAGATGAGCTGCCCGGCAAGTGAAGCCC 507  
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100  
QY 508 AAGATGATCGCGCGCATCGCGGCTTATCATAGGTGGCGCAGTACGACCGAGATCCTGATC 567  
Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleProVal 120  
QY 568 GAGATCTGCGCAGAGAGGCCATCGGCACCGTGTGATCGGCCCGCCCGCGTGAACATC 627  
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGGTGACACCTTGAACTTCCCATCAGCCCCATC 687  
Db 141 IleglyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
QY 688 GAGACCTGCGCGTGAAGCTGAAGCCCGGCATGACCGCCCAAGGTGAAGCAGTGCGCC 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
QY 748 CTGACCGAGGAGAAAGATCAAGGCGCTGACCGCCCATCTGCGAGGAGATGGAAAGAGGCGC 807  
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
QY 808 AAGATCAACCAAGATCGCGCGCGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAAG 867  
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
QY 868 AAGACACAGCAACCAAGTGGCGCAAGCTGGTGACCTTCGCGAGCTGAAACAAGCAGCCAG 927  
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
QY 928 GACTTCTGGGAGGTGACGTGGGCGATCCCCACCGCCCGCTGAAGAAAGAAAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
QY 988 GTGACCGTGTGAGCTGGGCGACGCTTCTTACAGCTGCCCTGGAGAGGACTTCCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
QY 1048 AAGTACACCGCTTACCATCCCGCAGCATCAACACGAGACCCCGCGCATCCGCTACCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
QY 1108 TACAACCTGTGCGCCCGAGGCTGGAAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTGAGCGCTTCCCGCGCCGCAACCCCGAGATCGTGATCTACCATCATGAGC 1227  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
QY 1228 GACTGTACCTGGCGAGCTGAGATCGCGCAGCAGCCCGCGCAGATCGAGGAGCTG 1287  
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1288 CGCAAGCACCTGTGCTGGCGCTTCCACACCCCGCAGCAAGAGACCAAGAGAGAGCC 1347  
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1348 CCTTCTCTGTGGATGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet 400  
QY 1408 CTGCGCGAGAGGAGAGCTGGAACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAC 1467  
Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1468 TGGCGCAGCAGATCTACTACCCCGCATCAAGGTGGCGCAGCTGTCAAGCTGCTCGCGGC 1527  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440  
QY 1528 GCCAAGCGCTGACCGACATCGTGCCTTACCGAGAGGCGCGAGCTGGAGCTGCGCGAG 1587  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
QY 1588 AACCGGAGATCTGCGCGAGCGCGCTGCACGGGTGTACTACGACCCCGCAGCAAGACCTG 1647  
Db 461 AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu 480  
QY 1648 GTGCGCGAGATCCAGAAAGCAGGCGCACGACAGTGGAGCTACACGATCTTACCGAGCGCC 1707  
Db 481 ValAlaGluIleGlnLysGlnGlyGlnThrTrpThrThrGlnIleThrGlnGluPro 500

QY	1708	TTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATCGCACCGCCACCAACAGCGTG	1767
Db	501	PhelyaenLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal	520
QY	1768	AAGACGTGACCGAGCGGTGCGAAGATCGCCATCGGAGAGCATCGTGATCTGGGCAAG	1827
Db	521	LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys	540
QY	1828	ACCCCAAGTTCGCTCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTAC	1887
Db	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpMetGluTyr	560
QY	1888	TGCGAGGCCACCTCGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTCGTGAAGCTG	1947
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu	580
QY	1948	TGTTACAGCTGAGAGAGAGCCCATCATCTCGGCGCGAGACCTTCTACGTGGAGCGGCC	2007
Db	581	TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla	600
QY	2008	GCCAAACCGGAGACCAAGATCCGCAAGCGCGCTACGTGACCGACCGGCGCGCAGAAG	2067
Db	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys	620
QY	2068	ATCGTGAGCTGACCGAGACCAACCAAGACCGAGCTGAGACCGCGGCGCGCAGAAG	2127
Db	621	ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla	640
QY	2128	CTCGAGGACCGGCGAGCGAGGTGAACATCGTGACCGACAGCGAGTACGCCCTCGGCATC	2187
Db	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle	660
QY	2188	ATCCNGCGCCAGCCGACAGAGCGGAGCGAGCTGTTGAACAGATCATCTGAGCAGCTG	2247
Db	661	IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu	680
QY	2248	ATCAAGAAGAGAGAGTGTACCTGAGCTGGGTGCGCGCCACCAAGGCGCATCGCGCGCAAC	2307
Db	681	IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn	700
QY	2308	GAGCAGATGACAGCTGTGTGAGCAGGCGCATCGGCAAGGTGCTGTTCTGGAGCGGCATC	2367
Db	701	GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle	720
QY	2368	GAT 2370	
Db	721	Asp 721	
RESULT 8			
GNVWVL			
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly			
C;Species: human immunodeficiency virus type 1, HIV-1			
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004			
C;Accession: A03967			
R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.			
Nature 313, 450-458, 1985			
A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi			
A;Reference number: A93355; MUID:85111157; PMID:2982104			
A;Accession: A03967			
A;Molecule type: DNA			
A;Residues: 1-1012 <MUE>			
A;Cross-references: UNIPROT:P03368			
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re			
C;Genetics:			
A;Gene: pol			
C;Superfamily: pol polyprotein			
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle			
F;66-164/Product: retropepsin #status predicted <RTP>			
F;90/Active site: Asp (shared with dimeric partner) #status experimental			
Alignment Scores:			
Pred. No.:	8.79e-153	Length:	1012

Score:	3477.00	Matches:	651
Percent Similarity:	94.22%	Conservative:	34
Best Local Similarity:	89.55%	Mismatches:	26
Query Match:	75.87%	Indels:	16
DB:	1	Gaps:	3
US-09-610-313B-30 (1-2469) x GNVWVL (1-1012)			
QY	238	GCCTTCCCTCCAGGCAAGCGCCGCGAGTTC-----	267
Db	4	AlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerPro	23
QY	268	-----CCAGCGAGCAGAACCCGCCAACAGCCACAGCCACCGAGCTGCAGGTG---	318
Db	24	ThrIleSerSerGluGlnThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrp	43
QY	319	---CGGGCGACNAACCCCGCAGCGAGCGCGCGCGCGAGCGCGCGCGCGCGCGCGT---	369
Db	44	GlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPhe	63
QY	370	AACCTTCCCGCAGATCACCTTGTGGCAGCGCCCTCGTGAGCATCAAGGTGGCGGCCAG	429
Db	64	AsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGlyGln	83
QY	430	ATCAAGAGGCGCTGTGGACACCGCGCGCGACACACCTGTCTGGAGGAGATGAGCCTG	489
Db	84	LeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeu	103
QY	490	CCCGGCAAGTGGAGCCCAAGATGATCGCGCGGATCGCGGGTTCATCAAGGTGCGCCAG	549
Db	104	ProGlyArgTrpLysProLysMetIleGlyGlyIleGlyPheIleLysValArgGln	123
QY	550	TACGACACAGATCTGTGAGATCTCGCGCAAGAGCGCATCGCACCGCTGTGATCGGC	609
Db	124	TyrAspGlnIleLeuIleGlnIleCysGlyHisLysAlaIleGlyThrValLeuValGly	143
QY	610	CCACACCCCGTGAACTCATCGCCCGCAACATGCTCAACCGAGCTGGGTGCAACCTGAAC	669
Db	144	ProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsn	163
QY	670	TTCCTCATGACCCCATGAGACCGTGCCTGTAAGCTGAAGCCCGGCGATGAGCGGCC	729
Db	164	PheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyPro	183
QY	730	AAGGTCAAGCAGTGGCCCTCGACCGAGGAGAAAGATCAAGGCCCTGACCGCATCTCGAG	789
Db	184	LysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThr	203
QY	790	GAGATGGAGAAGGAGGCAAGATCAACAAAGATCGCGCCCGAGAACCCCTTACACACCCC	849
Db	204	GluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrPro	223
QY	850	GTCTTCCCATCAGAAGAAGGACACACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAG	909
Db	224	ValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGlu	243
QY	910	CTCAACAAGCGCACCCAGACTTCTCGGAGGTCTGAGGTGCGCATCCCGACCCCGCGCG	969
Db	244	LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisLysProAlaGly	263
QY	970	CTGAAGAAGAAGACGCGTGCCTGTGAGCGTGGCGGACCGCTACTTTCAGCGTGCCC	1029
Db	264	LeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValPro	283
QY	1030	CTGGACGAGGACTTCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAGAGACC	1089
Db	284	LeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThr	303
QY	1090	CCCGGATCCGCTACCAAGTACACGCTGCTGCCCGCGGCTGGGAAGGCGACCCCGAGCATC	1149
Db	304	ProGlyLysArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIle	323
QY	1150	TTCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTG	1209

Db 324 PheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343  
Qy 1210 ATCTACAGTACATGGAGACCTGTACGTGGCGAGCGACCTGGAGATCGGCCACACCGC 1269  
Db 344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363  
Qy 1270 GCCAAGATCGAGGAGCTGGCGAAGACCTGTGGCTGGGGCTTCACACCCCGACAG 1329  
Db 364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLys 383  
Qy 1330 AAGCACCAGAGAGAGCCCTCTCTGTGGATGGCTACGAGCTGCCACCCGACAAGTGG 1389  
Db 384 LysHisGlnLysGluProPheLeuThrMetGlyTyrGluLeuHisProAspLysTrp 403  
Qy 1390 ACCGTGCGCCATCGAGCTGCCGAGAGAGCTGGACCGTGAACGACATCCAGAAG 1449  
Db 404 ThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLys 423  
Qy 1450 CTGTGGGCAAGCTGAATGGGCGAGCGACATCTACCCCGGCATCAAGTGGCGCAGCTG 1509  
Db 424 LeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443  
Qy 1510 TGCAGCTGTGGCGGCGCCCAAGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCC 1569  
Db 444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463  
Qy 1570 GAGCTGAGCTGGCGGAGAACCGGAGATCTCTGGCGAGCCCGTGCACGGCTGTACTAC 1629  
Db 464 GluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyr 483  
Qy 1630 GACCCCAAGAGACCTGTGGCGAGATCCAGAACGAGGCGCACGACCTGAGCTAC 1689  
Db 484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrpThrTyr 503  
Qy 1690 CAGATCTACAGGAGCCCTTCAAGACTGGAAGCCGCGAGTACGCAAGATCGGCACC 1749  
Db 504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523  
Qy 1750 GCCCACCAACGACGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATCGAGC 1809  
Db 524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543  
Qy 1810 ATCTGTATCTGGGCAAGACCCCAAGTTCGCTCGCCATCCAGAGAGACCTGGAG 1869  
Db 544 IleValIleTyrGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTrpGlu 563  
Qy 1870 ACCTGGTGGACGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACCC 1929  
Db 564 ThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThr 583  
Qy 1930 CCCCCTCTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCCGAGACC 1989  
Db 584 ProProLeuValLysLeuTyrTyrGlnLeuGlnLysGluProIleValGlyAlaGluThr 603  
Qy 1990 TTCTACGTGGACGCGCGCCCAACCGGAGACCAAGATCGCAAGCGCGCTACGTGACC 2049  
Db 604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLysAlaGlyTyrValThr 623  
Qy 2050 GACCGGCGCGGAGAGATCGTAGCTGACCGAGACCACCAACCGAGACCGAGCTG 2109  
Db 624 AsnLysGlyArgGlnLysValIleProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643  
Qy 2110 CAGGCCATCCAGCTGGCGCTCGAGACAGCGGAGCGAGGTGAACATCGTCAACCGACG 2169  
Db 644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAlaIleValThrAspSer 663  
Qy 2170 CAGTACGCCCTGGGCATCATCAGCGCCAGCCGACAAAGAGCGAGAGCGTGGTGAAC 2229  
Db 664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluLeuValAsn 683  
Qy 2230 CAGATCATCGAGGAGCTGATCAAGAGAGAGTGTACTGAGCTGGTCCCGCCAC 2289

Db 684 GlnIleIleGluGlnLeuLeuLysLysGluLysValTyrLeuAlaTrpValProAlaHis 703  
Qy 2290 AAGGCGATCCGCGCAACGACGACATCCACAAGCTGTGAGCAAGGGCATCCGCAAGTG 2349  
Db 704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723  
Qy 2350 CTGTTCTCTGGACGCGCATCGAT 2370  
Db 724 LeuPheLeuAspGlyIleAsp 730  
RESULT 9  
GNLJSI  
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz  
N:Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly.  
C:Species: simian immunodeficiency virus SIVcpz  
A:Host: Pan troglodytes (chimpanzee)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C:Accession: S09984  
R:Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.  
Nature 345, 356-359, 1990  
A:Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.  
A:Reference number: S09983; MUID:90259077; PMID:2188136  
A:Accession: S09984  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1027 <HUE>  
A:Cross-references: UNIPROT:P17283; EMBL:X52154  
F:81-180/Product: retropepsin #status predicted <RTP>  
F:105/Active site: Asp (shared with dimeric partner) #status predicted  
Alignment Scores:  
Pred. No.: 2,73e-142 Length: 1027  
Score: 3249.00 Matches: 599  
Percent Similarity: 89.46% Conservative: 63  
Best Local Similarity: 80.95% Mismatches: 75  
Query Match: 70.89% Indels: 3  
DB: 1 Gaps: 1  
US-09-610-313B-30 (1-2469) x GNLJSI (1-1027)  
Qy 159 CTGCTGGAAGTGGCGCAAGGAGGCGCACCATGAGGACTGCACCGAGCGCCAGGCCAA 218  
Db 7 LeuLeuAlaValTrpAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26  
Qy 219 CTCTCTCCGAGACCTGCGCTTCCCGAGGCGAAGCGCGAGTTCCCGACGAGCA 278  
Db 27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGlu 46  
Qy 279 GAACCGCGCCACAGCCCGACCGCGAGCTGCAGTGCAGTGCAGCGC-----GACACCC 332  
Db 46 nAsnArgThrAsnGlyProThrAspArgGluLeuTrpValProGlyGlyArgGluPr 66  
Qy 333 CCGCAGCGAGCG 392  
Db 66 oGlyGluGluArgGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86  
Qy 393 GCAGCGCCCTCTGTGAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452  
Db 86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106  
Qy 453 CGGCGCGCGACACCGCTGTGGAGGAGATGAGCTGCCCGCGCAAGTGGAGGCCACAGAT 512  
Db 106 rGlyAlaAspAspThrValIleGluArgIleGlnLeuGlnGlyLeuTrpLysProLysMe 126  
Qy 513 GATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 572  
Db 126 tIleGlyGlyIleGlyPheIleLysValLysGlnPheAsnValHisIleGluI 146



```
Qy 481 ATGAGCCTGCCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGGCTTCATCAAG 540
Db 1 MetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyPheIleLys 20
Qy 541 GTGGCCAGTACGACCAAGATCCTGATCGAGATCTCGCGCAAGAGGCGCATCGGCACCGTG 600
Db 21 ValLysGlnTrpAspGlnIleAlaIleIleGlyCysGlyHisLysAlaIleGlyThrVal 40
Qy 601 CTGATCGGCGCCCAACCCCGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGC 660
Db 41 LeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCys 60
Qy 661 ACCCTGAACCTCCCATCAGCCCATCGAGACCTGCGCGTGAAGCTGAAGCCCGGCATG 720
Db 61 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 80
Qy 721 GACGGCCCAAGGTGAAGTGAAGTGGCCCTGACCCAGGAGAGATCAAGGGCCCTACCGCC 780
Db 81 AspGlyProLysValLysGlnTrpProLeuThrGlnLysIleLysAlaLeuIleGlu 100
Qy 781 ATCTCGAGGAGATGAGAGAGGCGCAAGATCATCAAGATCGCGCCCGAGAACCCCTAC 840
Db 101 IleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyr 120
Qy 841 AACACCCCGTGTTCGCCATCAAGAGAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 121 AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp 140
Qy 901 TTCGGCAGCTGAACAGCGCACCCAGGACTTCTCGGAGGTGACGTGGGCATCCCCCAC 960
Db 141 PheArgGlnLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 160
Qy 961 CCCGCCGCTGAGAGAGAGAGAGCGTGCACCGTGTGGAGCTGGCGCGACCGCTACTTC 1020
Db 161 ProGlyGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 180
Qy 1021 AGCGTGGCCCTGAGAGAGACTTCGCAAGTACACCGCTTACCATCCCGAGCATCAAC 1080
Db 181 SerValProLeuAspLysAspPheArgLysTrpAlaPheThrIleProSerIleAsn 200
Qy 1081 AACGAGACCCCGCATCCGCTACAGTACAACTGCTGCCAGGCTGCCAGGCTGGAGGGCAGC 1140
Db 201 AsnGluThrProGlyIleArgTrpGlnTrpAsnValLeuProGlnGlnTrpLysGlySer 220
Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTCGAGCCCTTCCCGCGCCGCAACCCC 1200
Db 221 ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro 240
Qy 1201 GAGATCGTATCTACCAAGTACATGACGACTGTGACCTGGCGCGACGACCTGGAGATCGG 1260
Db 241 GluIleIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 260
Qy 1261 CAGCAGCGCCAGATCGAGGAGCTGCGCAAGCAGCTGCGCTGGGCTTCACCAAC 1320
Db 261 GlnHisArgThrLysIleGluGluLeuArgHisLysLeuLeuArgTrpGlyPheThr 280
Qy 1321 CCCGACAAGAGCACCAGAGGAGCCCTTCTGTGATGGGCTACGAGCTGCACCCC 1380
Db 281 ProAspLysLysHisGlnLysGluProPheLeuTrpIleGlyTyrGluLeuHisPro 300
Qy 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACCGTGAACGAC 1440
Db 301 AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp 320
Qy 1441 ATCCGAGAGCTGTGGCGAGCTGAAGTGGCCGACGAGATCTACCCCGGCATCAAGGTG 1500
Db 321 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal 340
Qy 1501 CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGCCCTGACCCAGCATCTGCGCCCTGACC 1560
Db 341 ArgGlnLeuCysArgLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr 360
Qy 1561 GAGGAGCGCAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGC 1620
```

```
Db 361 LysGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysThrProValHisGly 380
Qy 1621 GTGTACTACCAACCCAGAGGACCTGGTGGCCGAGATCCAGAGAGCGGGCCACGACCAG 1680
Db 381 ValTyrTrpAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyLeuGlyGln 400
Qy 1681 TGGACCTACCAAGTCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db 401 TrpThrTyrGlnIleTyrGlnProPheLysAsnLeuLysThrGlyLysTyrAlaLys 420
Qy 1741 ATGGCAGCCCGCCACACCAAGAGCTGACAGCTGACCGGCGCGTGCAGAAATCGCC 1800
Db 421 MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 440
Qy 1801 ATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGAG 1860
Db 441 ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu 460
Qy 1861 ACTTGGAGACCTGGTGGACCGACTACTTGGAGGCCACCTGGATCCCGAGTGGAGTTC 1920
Db 461 ThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 480
Qy 1921 GTGAACACCCCGCTGGTGAAGCTGTGTACAGCTGGAGTGGAGAGGCCCATCATCGGC 1980
Db 481 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 500
Qy 1981 GCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGC 2040
Db 501 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 520
Qy 2041 TACGTGACCGACCGGGCCCGGAGAGATCGTGAAGCTGACCGAGCACCAACAGAGAG 2100
Db 521 TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrAsnGlnLys 540
Qy 2101 ACCGAGCTGACGAGCATCCAGCTGGCCCTCGAGAGACGGCGACGAGGTGAACATCGTG 2160
Db 541 ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal 560
Qy 2161 ACCGACAGCCAGTACCGCTGGGCATCATCCAGGCCCGAGCCCGCAAGAGCGAGCGAG 2220
Db 561 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 580
Qy 2221 CTGGTGAACCAAGTATCGAGACGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTG 2280
Db 581 LeuValSerGlnIleIleGluHisLeuIleLysLysGluLysValTyrLeuAlaTrpVal 600
Qy 2281 CCCGCCACAGAGGCGCATCCGCGCAACGACAGATCGACAAGCTGGTGAAGCGGCGCATC 2340
Db 601 ProAlaHisLysGlyIleGlyGlyAsnGluValAspLysLeuValSerAlaGlyIle 620
Qy 2341 CGCAAGGTGCTGTCTCTGGACGCGCATCGAT 2370
Db 621 ArgLysValLeuPheLeuAspGlyIleAsp 630
```

## RESULT 11

T01668

pol polyprotein - human immunodeficiency virus type 1

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C;Accession: T01668

R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.

Cell 46, 63-74, 1986

A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol

A;Reference number: Z14389, MUID:86245056; PMID:2424612

A;Accession: T01668

A;Status: translated from GB/EMBL/DDBU

A;Molecule type: mRNA

A;Residues: 1-902 &lt;Ali&gt;

A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28012.1; PID:g60230

C;Superfamily: pol polyprotein

Alignment Scores:



Pred. No.:	6.48e-137	Length:	902
Score:	3132.00	Matches:	578
Percent Similarity:	97.74%	Conservative:	28
Best Local Similarity:	93.23%	Mismatches:	14
Query Match:	68.34%	Indels:	0
DB:	2	Gaps:	0

US-09-610-313B-30 (1-2469) x T01668 (1-902)

QY	511	ATGATCGGCGCATCGCGGCTTCATCAAGGTGGCGCAGTACGACCGAGTCTCTGATCGAG	570
Db	1	MetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu	20
QY	571	ATCTGGGCAAGAGGCCATCGCACCGTGTGATCGGCCCGCCACCCCGGTGAACATCATC	630
Db	21	IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle	40
QY	631	GGCGCAACATGTGACCCAGCTGGCTGCACCTGAACCTTCCCATCGCCCATCGAG	690
Db	41	GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	60
QY	691	ACCGTGCCTGAAGTGAAGCCCGCATCGACGGCCCGCCCAAGGTGAAGCATGGCCCTTG	750
Db	61	ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTyrProLeu	80
QY	751	ACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGTGGAGAGGGGCAAG	810
Db	81	ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys	100
QY	811	ATCACCAGAATCGGCCCGGAGAACCCCTACAACACCCCGCTTTCGCGCATCAAGAAAG	870
Db	101	IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys	120
QY	871	GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAGCGCACCCAGGAC	930
Db	121	AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp	140
QY	931	TTCTGGGAGTGGAGTGGCGATCCCGCCACCCCGCGCGCTGAAGAGAGAGAGCGTG	990
Db	141	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal	160
QY	991	ACCGTGTGGACGTGGCGCGCGCTACTTCAGCGTGGCCCTGGACGAGGACTTCCGCAAG	1050
Db	161	ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys	180
QY	1051	TACACGCCCTTACCATTCCCGAGCATCAACAAAGAGACCCCGCGCATCCGCTACGAGTAC	1110
Db	181	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr	200
QY	1111	AACGTGTGCTGCCAGGCGCTGMAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGCCAAG	1170
Db	201	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	220
QY	1171	ATCTGTGAGCCCTTCCCGCGCGCAACCCCGAGATCGTGTATCTACCATGATCGGACGAC	1230
Db	221	IleLeuGluProPheArgThrLysAsnProGluIleValIleIleTyrGlnTyrMetAspAsp	240
QY	1231	CTGTACTGGGCGACGACCTGGAGATCGGCAGCACCGCGCCCAAGATCGAGGAGCTGGCG	1290
Db	241	LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	260
QY	1291	AAGCACTGTGCTGGCGCTTCCACCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1350
Db	261	GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	280
QY	1351	TTCTGTGGATGGGCTACGAGCTGACCCCGCAAGGTGGACCGGTGACGCCCATCGAGCTG	1410
Db	281	PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu	300
QY	1411	CCGAGAGAGAGAGCTGGACCGGTGAACGACATCCAGAAAGCTGGTGGCGCAAGCTGG	1470
Db	301	ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	320

QY	1471	GCCAGCAGCATCTACCCCGGCATCAAGGTGGCGCAGTGTGCAAGCTGTGCGCGCGCC	1530
Db	321	AlaSerGlnIleIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla	340
QY	1531	AAGCCCTGACCGACATCGTGTCCCTTGACCGAGAGCGGAGGTGGAGTGGCGCGGAAC	1590
Db	341	LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn	360
QY	1591	CGCGAGATCTCGCGCGAGCCGCTGACCGGTGTACTAGACCCCGAGAGGACCTGGTG	1650
Db	361	ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle	380
QY	1651	GCCGAGATCCAGAAAGCGGCCACGACCGTGTGACCTACCGATCTACCGAGGACCCCTC	1710
Db	381	AlaGluIleGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluGlnTyr	400
QY	1711	AGAACTGAAGACCGGCAAGTACGCAAGATCGCAGATCGCACCGCCACCAACGACGTGAAG	1770
Db	401	LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys	420
QY	1771	CAGCTCAGCGAGCCGTCGAGAGATCGCCATCGAGAGATCGTGTGATCTGGGGCAAGACC	1830
Db	421	GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr	440
QY	1831	CCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG	1890
Db	441	ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp	460
QY	1891	CAGGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGG	1950
Db	461	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp	480
QY	1951	TACAGCTGGAGAGAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGGCC	2010
Db	481	TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla	500
QY	2011	AACCGGAGACCAAGATCGGCAAGCGCGCTAGCTGACCGACCGCGCGCGCGCAGAGATC	2070
Db	501	AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal	520
QY	2071	GTGAGCTGACCGAGACCCAAACCGAGAGACCGAGTGGAGCCATCGAGTGGCGCTG	2130
Db	521	ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu	540
QY	2131	CAGGACAGCGGCGAGCGAGTGAACATCGTACCGACGACGACGACGCGCTGGGCATCATC	2190
Db	541	GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle	560
QY	2191	CAGGCCAGCCCGACAGAGCGAGCGAGCTGGTGAACCCAGATCATCGAGCAGCTGATC	2250
Db	561	GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle	580
QY	2251	AGAAGAGAGAGTGTACTGTAGCTGGTGGTGGCGCCCGACCAAGGGCATCGCGCGCAAG	2310
Db	581	GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu	600
QY	2311	CAGATCGACAGCTGTGTGCAAGGCGCATCCGCAAGGTGTGTCTCTGGACGCGCATCGAT	2370
Db	601	GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp	620

RESULT 12

B47175  
reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1  
C;Species: human immunodeficiency virus type 1, HIV-1  
C;Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C;Accession: B47175  
R;Mohri, H.; Singh, W.T.; Ho, D.D.  
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993  
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the  
A;Reference number: A47175; MUID:93126353; PMID:7678340  
A;Accession: B47175  
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual trar  
A;Molecule type: DNA

A;Residues: 1-559 <MOH>

A;Cross-references: UNIPROT:Q9PXX1

A;Note: sequence extracted from NCBI backbone (NCBIP:122099)

C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.: 4,39e-122 Length: 559  
Score: 2809.00 Matches: 518  
Percent Similarity: 96.78% Conservative: 23  
Best Local Similarity: 92.67% Mismatches: 18  
Query Match: 61.29% Indels: 0  
DB: 2 Gaps: 0

US-09-610-313B-30 (1-2469) x B47175 (1-559)

```

Qy 673 CCATCAGCCCGCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATCGAGCGGCCCAAG 732
Db 1 ProleSerProleGluThrValProValValLysLeuLysProGlyMetAspGlyProLys 20
Qy 733 GTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGAGGAG 792
Db 21 ValLysGlnTrpProLeuThrGluGluLysLeuLysAlaLeuValGluLysCysThrGlu 40
Qy 793 ATGAGAGAGGAGGACAGATCACCAAGATCGGCCCGCGAGACCCCTACACACCCCGTG 852
Db 41 MetGluLysGluGlyLysLeuSerLysLeuGlyProGluAsnProTyrAsnThrProVal 60
Qy 853 TTGCCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGTGGACTTCCCGGAGCTG 912
Db 61 PheAlaLeuLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 80
Qy 913 AACAGCGCCACCAAGTCTTGGAGGTGAGCTGGGCATCCCGACCCCGCGGCTG 972
Db 81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeu 100
Qy 973 AAGNAGAGAGAGCGTGCCTGACCGTCTGAGCTGGCGGACCGCTACTTACGGTCCCGTG 1032
Db 101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120
Qy 1033 GACGAGACTTCCCAAGTACACCGCTTACCATCCCGACGATCAACACGACACCCCG 1092
Db 121 AspGluAspPheArgLysThrAlaPheThrLeuProSerLeuAsnGluThrPro 140
Qy 1093 GGATCCGCTACCAAGTACAAGTCTGCTCCCGAGGGCTGGAAGGCGACCCCGAGCATTC 1152
Db 141 GlyLeuArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaLeuPhe 160
Qy 1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATC 1212
Db 161 GlnSerSerMetThrLysLeuLeuGluProPheArgLysGlnAsnProAspLeuValle 180
Qy 1213 TACCAGTACATGAGCAGACCTGACGTGGGAGGACCTGGAGATCGGCGACCGCGCC 1272
Db 181 TyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluLysGlnHisArgThr 200
Qy 1273 AAGATCAGGAGCTGCCAAGACACCTCTCGCTGGGGCTTACACCCCGCGACAAGAG 1332
Db 201 LysLeuGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLys 220
Qy 1333 CACCAGAGAGGCCCCCTTCTGTGTGATGGGTACGAGCTGCACCCCGACAGTGGACC 1392
Db 221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240
Qy 1393 GTGCAGCCCATCGAGCTGCCGAGAAGGAGCTGGACCGTGAACGACATCCAGAACCTG 1452
Db 241 ValGlnProleValLeuProGluLysAspSerTrpThrValAsnAspLeuGlnLysLeu 260
Qy 1453 GTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGCGCGAGCTGC 1512
Db 261 ValGlyLysLeuAsnTrpAlaSerGlnLeuTyrProGlyLeuLysValArgGlnLeuCys 280
Qy 1513 AAGTGTCTGCGCGCGCAAGGCCCTGACCGACATCGTGCCTGACCGAGGAGCGCGAG 1572

```

```

Db 281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValleProLeuThrGluGluAlaGlu 300
Qy 1573 CTGAGCTGGCCGAGAACCCGCGAGATCTGCGCGAGCCCGTGACGGCGTGTACTACGAC 1632
Db 301 LeuGluLeuAlaGluAsnArgGluLeuLysGluProValHisGlyValTyrTrpAsp 320
Qy 1633 CCCAGCAAGACCTGTGGCGGAGATCCAGAGCAGGCGCCAGCAGCAGTGCCTACCCAG 1692
Db 321 ProSerLysAspLeuLeuAlaGluLeuGlnLysGlnGlyGlnTrpThrTyrGln 340
Qy 1693 ATCTTACCAGAGCCCTTCAAGAACCTGAAGACCCGCAAGTACGCGCAAGATCGGCAACGCC 1752
Db 341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLeuMetArgGlyAla 360
Qy 1753 CACACCAACGCTGAAGCAGCTGACCGAGGCCGTGAGAGATCGCCATCGAGAGCATC 1812
Db 361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerile 380
Qy 1813 GTGATCTGGGCAAGACCCCAAGTTCGGCTCGCCATCCAGAGGAGACCTGGGAGACC 1872
Db 381 ValIleTrpGlyLysThrProArgPheLysLeuProIleGlnLysGluThrTrpGluThr 400
Qy 1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGCGAACACCCCC 1932
Db 401 TrpTrpThrGluTyrTrpGlnAlaThrTrpLeuProGluTrpGluPheValAsnThrPro 420
Qy 1933 CCCCTGTGAAGCTGTGTATACAGCTGAGAGAGAGCCCATCATCGGCGCGAGACCTTC 1992
Db 421 ProLeuValLysLeuTrpTyrGlnLeuLysGluProIleValGlyAlaGluThrPhe 440
Qy 1993 TAGCTGACCGCGCGCCCAACCCGCGAGACCAAGATCGGCAAGGCGGCTAGTGCACCGAC 2052
Db 441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460
Qy 2053 CGGGCGCGGAGAGATCGTGACCTGACCGAGACCAACCAAGAGAGACCGAGCTGCAG 2112
Db 461 LysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeuGln 480
Qy 2113 GCATCCAGCTGGCCCTGCGAGGACGCGGAGCGAGGTGAACATCGTGCACCGAGCCAG 2172
Db 481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500
Qy 2173 TAGCCCTGGGCATCATCAGGCCCGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACCGAC 2232
Db 501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520
Qy 2233 ATCATCGAGCAGCTGTCAAGAGGAGAGTGTACCTGAGCTGGTGGCGCGCCACCAAG 2292
Db 521 IleIleGlnGlnLeuIleLysLysValTyrLeuAlaTrpValProAlaHisLys 540
Qy 2293 GGATCCGCGCCCAACGAGCAGATCGACAAGCTGTGAGCAAGGGCGCATCCCGCAAGGTG 2349
Db 541 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559

```

# RESULT 13

A47175

reverse transcriptase, AZT-resistant variant - human immunodeficiency virus type 1

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004

C;Accession: A47175

R;Mohr, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.

Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993

A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the

A;Reference number: A47175; PMID:93126353; PMID:7678340

A;Accession: A47175

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra

A;Molecule type: nucleic acid

A;Residues: 1-559 <MOH>

A;Cross-references: UNIPROT:Q9PXX2

A;Note: sequence extracted from NCBI backbone (NCBIP:122100)

C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.: 3.66e-121 Length: 559  
 Score: 2789.00 Matches: 514  
 Percent Similarity: 96.60% Conservatives: 26  
 Best Local Similarity: 91.95% Mismatches: 19  
 Query Match: 60.86% Indels: 0  
 DB: 2 Gaps: 0

US-09-610-313B-30 (1-2469) x A47175 (1-559)

QY	673	CCCATCAGCCCATTCGAGACCGTCCCGTGAAGCTGAAGCCCGGATGAGCGGCCCAAG	732
Db	1	ProileSerProilleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	20
QY	733	GTCAAGCAGGCGCCCTGACCGAGGAGAAGATCAAGSCCTGACCGCATCTCGGAGGAG	792
Db	21	ValLysGlnTrpProLeuThrGluLysLeuValLysLeuValGluLysCysThrGlu	40
QY	793	ATGAGAAAGAGGCGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACAAACCCCGGTG	852
Db	41	MetGluLysGluGlyLysIleSerLysIleGlyProGluAenProTyrAenThrProVal	60
QY	853	TTGCCATCAAGAAAGAGCAGACCAAGTGGCGCAAGCTGGTGACTTCCGCGAGCTG	912
Db	61	PheAlaIleLysLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu	80
QY	913	AACAAGCGCACCCAGGACTTCTGGAGGTGACGTGGGCATCCCGACCCCGCGGCTG	972
Db	81	AsnLysArgThrGlnAspPheTrpGluValGlnLysGlyIleProHisProAlaGlyLeu	100
QY	973	AAGAAGAAAGAGCGTGCCTGCTGGACGTGGCGACGCGCTACTTACGCTGCGCCCTG	1032
Db	101	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	120
QY	1033	GAGAGGACTTCGCGAAGTACACGCGCTTCCACATCCCGAGCATCAACAGAGACCCCG	1092
Db	121	AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAenAenGluThrPro	140
QY	1093	GGATCCGCTACCAAGTACACGCTGCTCCCGAGGCTGGAAGGCGACGCCCGACCTTC	1152
Db	141	GlyLeuArgTyrGlnTyrAenValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	160
QY	1153	CAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC	1212
Db	161	GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAenProAspMetValIle	180
QY	1213	TACAGTACATGACGACCTGTAGCTGGGAGGACCTGAGAGATCGCGCAGCAGCGGCC	1272
Db	181	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr	200
QY	1273	AAGATCGAGGAGCTGCGCAAGCACCTGCTCGCTGGGGCTTCAACACCCCGCAAGAAG	1332
Db	201	LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys	220
QY	1333	CACCAAGAGGAGCCCTTCTGTGTGATGGGTACGAGCTGCACCCCGCAAGTGGACC	1392
Db	221	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	240
QY	1393	GTGAGCCCATCGAGCTGCGGAGAGAGAGCTGGACCTGAGCATCGACATCCAGAGCTG	1452
Db	241	ValGlnProIleValLeuProGluLysAspSerTrpThrValAenAspIleGlnLysLeu	260
QY	1453	GTGGGCAAGCTGAACCTGGCGCAGCATCTTACCCCGGCATCAAGGTGCGCGCTGTC	1512
Db	261	ValGlyLysLeuAenTrpAlaSerGlnIleTyrProGlyLysValLysGlnLysCys	280
QY	1513	AAGTGTCTGGCGCGCAAGCCCTGACCGACATCGTCCCTGACCGGAGGCGCGAG	1572
Db	281	LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluGluAlaGlu	300
QY	1573	CTGAGCTGGCGGAGAACCGGAGATCTCTGGCGAGCCGTCGACCGCGTGTACTAGAC	1632
Db	301	LeuGluLeuAlaGluAenArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp	320

QY	1633	CCCAAGAGGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCGAGTGCACCTACCAG	1692
Db	321	ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnTyrThrTyrGln	340
QY	1693	ATCTACAGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGGACCGCC	1752
Db	341	IleTyrGlnGluProPheLysAenLeuLysThrGlyLysTyrAlaArgMetArgGlyAla	360
QY	1753	CACACCAAGCAGCTGAAGCAGCTGACCGAGGCGCTCAGAAAGATCGCATGGAGAGCATC	1812
Db	361	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle	380
QY	1813	GTGATCTGGGCGCAAGACCCCAAGTTCCGCTCCCTCCATCCAGAAGGAGACCTGGAGACC	1872
Db	381	ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla	400
QY	1873	TGTTGACCGCATCTACCTGGCAGGCCACTGTGATCCCGAGTGGGAGTTCGTGAACACCCC	1932
Db	401	TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAenThrPro	420
QY	1933	CCCTCGTGAAGCTGTGTACCAAGCTGGAGAGGCCATCATCGCGCGGAGACCTTC	1992
Db	421	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe	440
QY	1993	TAGCTGGACGCGGCCCAACCGCAGAGACCAAGATCGGCAAGGCCGCTACGTACACGAC	2052
Db	441	TyrValAspGlyAlaAlaAenArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	460
QY	2053	CGGGCGCGCAGAAAGTCTGTGACCTGACCGAGACCAACCAACAGAGACCGAGCTGCAG	2112
Db	461	LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln	480
QY	2113	GCCATCCAGCTGCGCCCTGCAGGACAGCGCAGCGAGGTGAACATCGTACCGCAGCGCAG	2172
Db	481	AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAenIleValThrAspSerGln	500
QY	2173	TACGCGCTGGGCATCATCCAGGCCCGCAGCAGCAGAGCGAGCGAGCTGGTGAACACG	2232
Db	501	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln	520
QY	2233	ATCATCGACGAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGGTGCGGCCCAAG	2292
Db	521	IleIleGluLeuLeuLysLysGlyLysValTyrLeuAlaTrpValProAlaHisLys	540
QY	2293	GGCATCGGCGGCAACGAGCAGATCGCAAGCTGCTGAGCAAGGCGCATCCCGAAGGTG	2349
Db	541	GlyIleGlyLysAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal	559

RESULT 14

S46347  
 pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)  
 C;Species: simian immunodeficiency virus SIVagm  
 A;Variety: isolate SAB-1  
 C;Date: 25-Dec-1994 #sequence\_revision 14-Feb-1997 #text\_change 26-Aug-1999  
 C;Accession: S46347  
 R;Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.; et al.  
 EMBO J. 13, 2935-2947, 1994  
 A;Title: Mosaic genome structure of simian immunodeficiency virus from West African green  
 A;Reference number: S46335; MUID:94298785; PMID:8026477  
 A;Accession: S46347  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-1039 <JIN>  
 A;Cross-references: EMBL:U04005; NID:9466229; PIDN:AAA21505.1; PID:9466231  
 A;Experimental source: isolate SAB-1; sabaeus monkey  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993  
 C;Genetics:  
 C;Gene: pol  
 C;Superfamily: pol polyprotein  
 Alignment Scores: 1.13e-102 Length: 1039  
 Pred. No.:

Score: 2388.00 Matches: 444  
Percent Similarity: 74.60% Conservative: 120  
Best Local Similarity: 58.73% Mismatches: 150  
Query Match: 52.11% Indels: 42  
DB: 2 Gaps: 6

US-09-610-313B-30 (1-2469) x S46347 (1-1039)

QY 220 TTCCTCCGAGGAGCTGGCTTCCCGAGGCG-----AAGCGCGGAGTTCCCGAGC 273  
Db 1 PhePheArg-----ValTrpProLeuGlnArgGluThrGlnGluPheProSer 17  
QY 274 GAGCAGAACCGCGCCAAAC---AGCCCCACCAGCGCGGAGCTGCAGTGGCGGC----- 324  
Db 18 AspleuHisGlnThrAsnSerSerProAsnGlyThrGlyLeuGlnGlnAlaGlyGlyLys 37  
QY 325 -----GACAAACCCCGCAGCGAGGCG----- 345  
Db 38 LeuValCysArgGlnThrSerAspGlnArgThrArgAlaArgSerSerAsnSerPro 57  
QY 346 -----GGCGCGGAGCGC----- 357  
Db 58 VallysAlaValCysSerGlyGluThrAlaGluThrAlaValAlaLysProLeuAla 77  
QY 358 -----CAGGGACCTGAACTTCCCGAGATCACCCTGTGGCAGCGCCCC 402  
Db 78 ThrThrGluProLeuArgGlyGlyLeuGlnLeuProGlnValSerLeuTrpArgArgPro 97  
QY 403 CTGTGTGACATCAAGTGGCGGCGGCGAGATCAAGGAGGCGCTGTGCGACACCGCGCGGAC 462  
Db 98 MetLysThrValTrpIleGluGlyGlnLysValThrAlaLeuLeuAspThrGlyAlaAsp 117  
QY 463 GACACCGTGTGGAGAGATGAGCTGCCCGCAAGTGGAGCCCAAGATGATCGCGCGC 522  
Db 118 AspSerValIleGlnGlyIleGluLeuGlyAspAsnTrpLysProArgIleIleGlyGly 137  
QY 523 ATCGCGGCTTCATCAAGTGGCGGCGGAGTACGACCGAGATCCTGATCGAGATCTGGCGAAG 582  
Db 138 IleGlyGlyCysIleAsnValLysAlaTrpHisAsnGlnGluValLysIleGluAspLys 157  
QY 583 AAGGCCATCGGACCGCTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 642  
Db 158 ThrCysLysAlaThrIleLeuValGlyGluThrProValAsnIleIleGlyArgAsnVal 177  
QY 643 CTGACCCAGCTGGGTGACCTGAACTTCCCGATCAGCCCGCTGAGACCGTGGCGGTG 702  
Db 178 LeuAlaGlnLeuGlyValThrLeuAsnLeuThrGlnArgGluIleGluProIleLysVal 197  
QY 703 AAGCTGAAGCCCGGCGGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 762  
Db 198 HisLeuLysProGlyGlnAspGlyProArgIleArgGlnTrpProLeuSerLysGluLys 217  
QY 763 ATCAAGGCGCTGACCGGCTGCGGAGGAGATGAGAGAGAGGCGGCGGCGGCGGCGGCGG 822  
Db 218 IleGluAlaLeuLysAlaIleCysGluAspLeuGluLysGlnGlyHisLeuGluArgIle 237  
QY 823 GGCCCCGAGAACCCCTACAACACCCCGTGTGGCCATCAAGAGAGAGGAGGAGGAGGAGG 882  
Db 238 GlyProGluAsnProTrpAsnThrProValPheAlaIleArgLysLysAspLysThrGln 257  
QY 883 TGGGCGAAGCTGTGGAGTCTCGGAGTGAACAGCGCGACCCAGGATCTTGGGAGGTG 942  
Db 258 TrpArgIleLeuMetAspPheArgGlnLeuAsnLysSerThrGlnAspPheGlnGluVal 277  
QY 943 CAGTGGGCGATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1002  
Db 278 GlnLeuGlyIleProHisProAlaGlyLeuGlnArgGluGlnIleThrValLeuAsp 297  
QY 1003 GTGGCGGCGGCTTACTTCAGGCGTGGCTGGAGGAGGAGTTCGCGAAGTACACCGGCTTC 1062  
Db 298 IleGlyAspAlaTrpPheSerCysProLeuAspProAspPheGlnLysTrpAlaPhe 317  
QY 1063 ACCATCCCGGAGATCAACAGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1122

Db 318 ThrIleProSerValAsnAsnArgGluProGlyIleArgTrpGlnTrpLysValLeuPro 337  
QY 1123 CAGGCTCGAAGGCGGAGCCAGCATCTTCCAGAGCAGCATCACCAAGATCTCTGAGCGCC 1182  
Db 338 GlnGlyTrpLysGlySerProThrIlePheGlnThrThrAlaAsnLysIleLeuGlnGlu 357  
QY 1183 TTCGCGCGCGCAACCCCGAGATCGTATCTACAGTATCATGAGCAGCATCTGACGTGGCG 1242  
Db 358 PheArgGlnLysAsnProAspValAspIleTrpGlnTrpMetAspAspMetLeuIleAla 377  
QY 1243 AGCGACCTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1302  
Db 378 SerAspArgProLysAlaGluHisLeuValMetValGlnGlnLeuArgAspTrpLeuGlu 397  
QY 1303 CGCTGGGCTTCCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1362  
Db 398 ThrTrpGlyPheLysThrProGluLysLysPheGlnLysAspProProTrpLeuTrpMet 417  
QY 1363 GGCTACGAGTGCACCCCGGAGCAAGTGGACCGTGCAGCCCATCGACTGCCGAGAGGAG 1422  
Db 418 GlyTrpGluLeuTrpProLysLysTrpGlnLeuGlnGluIleThrLeuProGluArgGlu 437  
QY 1423 AGCTGGACCGTGAACGACATCCAGAGCTGTGGCGGCGGCGGCGGCGGCGGCGGCGG 1482  
Db 438 GluTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 457  
QY 1483 TACCGCGGCGATCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1542  
Db 458 TyrThrGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyAlaArgProLeuThr 477  
QY 1543 GACATCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1602  
Db 478 GluIleValGlnTrpThrGluGluAlaGluLeuGluLeuGluLysArgGlnIleLeu 497  
QY 1603 CGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1662  
Db 498 ArgGlnLysGlnGlnGlyGlnTrpTyrAspProAlaLeuProLeuArgAlaLysValLeu 517  
QY 1663 AAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1722  
Db 518 LysLeuGlyAspGlyGlnTrpGlyTrpGlnIleTrpGlnProGluLysLysIleLeuLys 537  
QY 1723 ACCGCGAGTACGCCAAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1782  
Db 538 ValGlyLysTrpAlaLysIleLysThrAlaHisThrAsnGluLeuArgMetLeuAlaGly 557  
QY 1783 GCGTGCAGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1842  
Db 558 LeuValGlnLysIleGlyLysGluSerIleValIleTrpGlyGlnIleProIleMetGlu 577  
QY 1843 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1902  
Db 578 LeuProValGluArgGluLeuTrpGluGlnTrpSerAspTrpTrpGlnValThrTrp 597  
QY 1903 ATCCCGGAGTGGGAGTTCGTAACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1962  
Db 598 IleProGluTrpGluMetValSerThrProGlnLeuIleArgLeuTrpTrpLysLeuVal 617  
QY 1963 AAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2022  
Db 618 LysAspProIleProGlyGluAlaValTrpValAspGlyAlaAlaAsnArgAsnSer 637  
QY 2023 AAGTTCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2082  
Db 638 LysGluGlyLysAlaGlyTrpLeuThrAspArgGlyAspGlnLysValValAlaLeuGlu 657  
QY 2083 GAGACCCACCAAGAGACCGGAGCTGCAGGCCATCCAGCTGGCGGCGGCGGCGGCGGCGG 2142  
Db 658 AsnThrThrAsnGlnLysAlaGluLeuGluAlaIleLeuLeuAlaLeuArgAspSerGly 677  
QY 2143 AGCGAGGTGAACATCTGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2202

Db 678 SerLysValAsnIleThrAspSerGlnTyraLaMetGlyIleAlaGlyGluPro 697  
QY 2203 GACAAGAGCAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAAG 2262  
Db 698 ThrGluSerAspAenAenIleValGlnGlnIleGluLeuIleIleLysGluAla 717  
QY 2263 GTGTACCTGAGCTGGTGGCCGCCCAACAAGGGCATCGCGCGCAACGAGCAGATCGACAAG 2322  
Db 718 ValTyriLeAlaTrpValProAlaHisLysGlyValGlyGlyAenGluGluIleAspLys 737  
QY 2323 CTGGTGAAGAGGATCCCAAGGTGCTGTTCTGGACGGCATCGAT 2370  
Db 738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753

RESULT 15  
S53092  
pol polyprotein - human immunodeficiency virus type 2  
C;Species: human immunodeficiency virus type 2, HIV-2  
C;Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
C;Accession: S53092  
R;Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.  
submitted to the EMBL Data Library, March 1995  
A;Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu  
A;Reference number: S53091  
A;Accession: S53092  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1055 <REC>  
A;Cross-references: UNIPROT:O73194; EMBL:Z48731  
C;Superfamily: pol polyprotein  
C;Keywords: polyprotein

Alignment Scores:  
Pred. No.: 4.8e-96 Length: 1055  
Score: 2244.00 Matches: 421  
Percent Similarity: 71.11% Conservative: 128  
Best Local Similarity: 54.53% Mismatches: 184  
Query Match: 48.96% Indels: 39  
DB: 2 Gaps: 9

US-09-610-313B-30 (1-2469) x S53092 (1-1055)

QY 156 GGGTGTCTGAAGTGGCGGCAAGGAGGGCCACCAGATGAAGGACTGCAC----CGAGCGCCA 212  
Db 3 GlyLeuLeuGluMetTrpGln-----AspArgThrTyHisGlyLysValPro 18  
QY 213 GGCCAACTT-----CTCCG 227  
Db 19 ArgGlnThrGlyGlyPhePheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38  
QY 228 CGAGGACCTGGCTTCCCCAGGGCAAGGCCCGAGTTCCTCCAGCGAGCAGAA----- 281  
Db 39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerGly 58  
QY 282 CCGCGCAACAGCCCAACAGCCGCGCA-----GCTGCAGGTGGCGCGCACCAACCCCGC 335  
Db 59 ProThrGlyGluIleTyraAlaArgGluLysAlaGluAlaGluArgGluThrIle 78  
QY 336 CAGCAGCGCGCGC-----CGAGCGCCA-----GGGCACCTCGAACTT----- 374  
Db 79 GlnArgGlyAspArgGlyLeuAlaAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98  
QY 375 -----CCCCAGATCACCTGTGCAGCGCCCTCGTGTGAGCATCAA 416  
Db 99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValThrAlaHi 118  
QY 417 GGTGGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGACACACCGTGTCTGGA 476  
Db 118 sIleGluGlyGlnProValGluValLeuLeuAspThrGlyAlaAspSerIleValAl 138  
QY 477 GGAGATAGCGCTGCCCGGCAAGTGAAGCCAGATGATCGCGGCGATCGCGCGCTTCAT 536  
Db 138 agIyIleGluLeuGlySerAsnTyrsrProLysIleValGlyIleGlyPheI 158

QY 537 CAAGGTGCGCCAGTACGACCCAGATCCTGATCGAGATCTGCGCAAGAAGCCATCGGCAC 596  
Db 158 eAsnThrLysGluTyriLysAsnValGluIleGluValLeuGlyLysArgValargAlaTh 178  
QY 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGG 656  
Db 178 rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuG 198  
QY 657 CTGCACCTCGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGG 716  
Db 198 yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProG 218  
QY 717 CATGACGCGCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAAGATCAAGGCCCTGAC 776  
Db 218 yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeu 238  
QY 777 CGCATCTCGAGAGATGAGAGGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCC 836  
Db 238 sGluIleCysGluLysMetGluArgGluGlyGlnLeuGluAlaProProThrAsnPr 258  
QY 837 CTACAACACCCCGTCTTCCCATCAAGAGAGGAGCAGACCAAGTGGCGCAAGCTGGT 896  
Db 258 ofrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysTrpArgMetLeu 278  
QY 897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGCGCATCCC 956  
Db 278 eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyIlePr 298  
QY 957 CCACCCCGCGCGCTGAAGAAGAGAGCGTGACCGTCTCGAGCTGGCGGACGCGCTA 1016  
Db 298 oHisProAlaGlyLeuAlaLysLysArgIleThrValLeuAspValGlyAspAlaTy 318  
QY 1017 CTTACGGTGCCTCGAGGAGACTTCCGCAAGTACACCGCTTCCACCATCCCCAGCAT 1076  
Db 318 rPheSerIleProLeuHisGluAspPheArgGlnTyThrAlaPheThrLeuProSerVa 338  
QY 1077 CAACAACGAGACCCCGCGCATCCGCTACCAAGTCAACAGCTGCTGCCCGAGGGTGGAGGG 1136  
Db 338 lAsnAsnAlaGluProGlyLysArgTyriLysValLeuProGlnGlyTrpLysG 358  
QY 1137 CAGCCCGAGATCTTCAGAGCAGCATGACCAAGATCTGGAGCCCTTCCGCGCGCCGCAA 1196  
Db 358 ySerProAlaIlePheGlnTyThrMetArgGlnValLeuGluProPheArgLysAlaAs 378  
QY 1197 CCGCGAGATCGTATCATCCAGTACATGAGCAGCTGTACGTGGCGAGCGACCTGGAGAT 1256  
Db 378 nGlnAspValIleIleGlnTyThrMetAspIleLeuIleAlaSerAspArgThrAs 398  
QY 1257 CGGCCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGCTTCAC 1316  
Db 398 pLeuGluHisAspArgValValLeuGlnLeuLysGluLeuLeuAsnSerLeuGlyPheSe 418  
QY 1317 CACCCCGCAGACAAGACACAGAGAGCCCGCTTCTGATGGTGTGATGGCTACGAGCTGCA 1376  
Db 418 rThrProAspGluLysPheGlnLysAspProProTyriArgTrpMetGlyTyriGluLeuTr 438  
QY 1377 CCGCGACAGTGGACCGCTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACCGTAA 1436  
Db 438 pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysGluValTrpThrValAs 458  
QY 1437 CGACATCCAGAAGCTGTGGCAAGCTGAACCTGGCGCAGCCAGATCTTACCCCGGATCAA 1496  
Db 458 nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyriProGlyIleLy 478  
QY 1497 GTGCGCCAGCTGTGAAGCTGTGCGCGCGCGCAAGGCCCTGACCGACATCGTGGCCCT 1556  
Db 478 sThrLysHisLeuCysArgLeuIleArgGlyLysMetThrLeuThrGluGluIleGlnTr 498  
QY 1557 GACCGAGGAGCGCGCTGGAGCTGCCGAGAACCGGAGATCTTCCGCGGACCGCGTGA 1616  
Db 498 pThrGluLeuAlaGluLeuGluAsnArgValIleLeuSerGlnGluGlnG 518





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 11:33:41 ; Search time 168.242 Seconds  
(without alignments)  
11351.628 Million cell updates/sec

Title: US-09-610-313B-30  
Perfect score: 4583  
Sequence: 1 gtagacccaccatggcgca.....gggctagaccggtgaattc 2469

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2 1/USPTO spool/US09610313/runat 31052005 15136 15139/app query.fasta\_1.7893  
-DB=A Geneseq 16Dec04 -OPMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CGN 1 1 557 @runat 31052005 15136 15139 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6  
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 16Dec04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3804.5	83.0	1435	ADS19486	ADS19486 Modified
2	3782.5	82.5	3183	ADP84803	ADP84803 HIV-1 hyb
3	3741.5	81.6	1457	ADN36406	ADN36406 Human pro
4	3658	79.8	1003	AAB69289	AAB69289 HIV-1 non
5	3649	79.6	999	3 AAB69286	AAB69286 HIV-1 non
6	3644	79.5	1005	3 AAB69287	AAB69287 HIV-1 non
7	3608	78.7	998	6 AAE37601	AAE37601 HIV-1 sub
8	3601	78.6	3025	4 AAB86169	AAB86169 HIV-1 sub
9	3555.5	77.6	1000	3 AAB69282	AAB69282 HIV-1 non
10	3537.5	77.2	1002	2 AAW72993	AAW72993 HIV isola

11	3537	77.2	854	5	AAM48949	Aam48949 HIV-1 sub
12	3535	77.1	1003	6	AAO30963	Aao30963 HIV pol p
13	3535	77.1	1003	8	ADN36413	Adn36413 HIV prote
14	3532.5	77.1	1002	1	AAP81861	Aap81861 Sequence
15	3532	77.1	1003	6	ABR55489	Ab-55489 Amino aci
16	3530	77.0	1003	1	AAP60420	Aap60420 Sequence
17	3530	77.0	1003	1	AAP70861	Aap70861 Sequence
18	3524	76.9	1003	3	AAI70602	Aay70602 Codon opt
19	3524	76.9	1003	3	AAI70601	Aay70601 Corrected
20	3522	76.8	1003	5	AAO19387	Aao19387 Lymphaden
21	3516.5	76.7	1003	5	AAO19387	Aao19387 Lymphaden
22	3513	76.7	1015	1	AAP60347	Aap60347 HIV-1 pol
23	3513	76.7	1015	2	AAR43867	Aar43867 HTLV-III
24	3513	76.7	1015	4	AAB85993	Aab85993 Amino aci
25	3510	76.6	1003	3	AAB10047	Aab10047 HIV-1 pol
26	3510	76.6	1003	3	AAI70600	Aay70600 Wild type
27	3509	76.6	1016	2	AAR08062	Aar08062 AcNPV-HIV
28	3506	76.5	1015	2	AAR43875	Aar43875 HTLV-III
29	3502	76.4	1016	2	AAR08063	Aar08063 HIV-1 pol
30	3499	76.3	2033	2	AAR08056	Aar08056 HIV-1 pol
31	3495	76.3	1016	2	AAR08054	Aar08054 HIV-1 pol
32	3493	76.2	1003	3	AAB69284	Aab69284 HIV-1 non
33	3492	76.2	2033	2	AAR08055	Aar08055 HIV-1 pol
34	3491	76.2	1003	8	ADP20076	Adp20076 Human imm
35	3487	76.1	1003	1	AAP61508	Aap61508 Sequence
36	3487	76.1	1003	2	AAR29705	Aar29705 pol gene
37	3487	76.1	1003	3	AAI77301	Aay77301 HIV-1 (AT
38	3487	76.1	1003	5	AAE35789	Aae35789 ARV-2 (9B
39	3487	76.1	1003	6	ABU63185	Abu63185 Pol prote
40	3487	76.1	1491	1	AAP91048	Aap91048 Transcrip
41	3484	76.0	1003	3	AAB69279	Aab69279 HIV-1 non
42	3482	76.0	739	5	AAU11874	Aau11874 HIV pol p
43	3482	76.0	1003	6	ABU57552	Abu57552 AIDS asso
44	3480	75.9	1012	6	ABU63325	Abu63325 Human lym
45	3477.5	75.9	1003	2	AAR08059	Aar08059 HIV-1 pol

#### ALIGNMENTS

RESULT 1  
ADS19486  
ID ADS19486 standard; protein; 1435 AA.  
XX  
AC ADS19486;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Modified HIV-1 GagPol protein.  
XX  
KW Gene delivery; epithelial cell; respiratory system; mammal;  
KW lentiviral packaging system; expression vector; gagpol; gene expression;  
KW lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;  
KW cystic fibrosis; CF; deltaPD; transmembrane conductance regulator; CFTR;  
KW gene therapy; HIV-1; mutant; mutein.  
XX  
OS Human immunodeficiency virus 1.  
OS Synthetic.  
XX  
PN US2004037780-A1.  
XX  
PD 26-FEB-2004.  
XX  
PF 23-AUG-2002; 2002US-00226638.  
XX  
PR 19-NOV-2001; 2001AU-00008942.  
XX  
PA (PARS)/ PARSONS D.  
PA (ANSO)/ ANSON D.  
PA (LIMB)/ LIMBERIS M.  
PA (FULL)/ FULLER M.  
XX  
PI Parsons D, Anson D, Limberis M, Fuller M;  
XX



Db 783 GlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGlyPheThr 802  
QY ACCCCGCAAGAAGCACCAGAGAGCCCTCTCTGTGTGGTGGGTACGAGCTGCAC 1377  
Db 803 ThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyThrGluLeuHis 822  
QY CCGGCAAGTGGACCGTGCAGCCCATCGAGTGCCTCCGAGAGAGAGAGTGCAGCGTGAAC 1437  
Db 823 ProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsn 842  
QY GACATCCAGAGCTGTGGCAAGCTGAACCTGGCCAGCCAGATCTACCCGGGATCAAG 1497  
Db 843 AspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyAlaGlyIleLys 862  
QY GTGGCCAGCTGTGCAAGCTGTGGCGCGCCAGCGCCCTCAGCATGTGCCCCCTG 1557  
Db 863 ValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeu 882  
QY ACCGAGAGGCGGAGCTGGAGCTGCCGAGAACCGCGAGATCTCGCGGAGCCCGTGCAC 1617  
Db 883 ThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis 902  
QY GCGGTGTACTAGCACCACGAGACCTGTGGCCGAGATCCAGAGACGAGGCGCACGAC 1677  
Db 903 GlyValTyThrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGly 922  
QY CAGTGGACCTACCATCTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCC 1737  
Db 923 GlnTrpThrTyThrGlnIleTyThrGlnProPheLysAsnLeuLysThrGlyLysTyAla 942  
QY AAGATGGCCACCGCCCAACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATC 1797  
Db 943 ArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIle 962  
QY GCCATGGAGACATCGTGTATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAG 1857  
Db 963 AlaThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLys 982  
QY GAGACCTGGAGACTGTGTGACGACCTACTGCGAGGCGACCTGATCCCGAGTGGGAG 1917  
Db 983 GluThrTrpGluThrTrpTrpThrGluTyThrGlnAlaThrTrpIleProGluTrpGlu 1002  
QY TTCGTGAACACCCCGCTCGTGAAGCTGTGGTACCAAGTGGAGAGGACCCCATCATC 1977  
Db 1003 PheValAsnThrProLeuValLysLeuTrpTyThrGlnLeuGluLysGluProIleIle 1022  
QY GCGCGGAGACTTCTAGCTGACGCGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCC 2037  
Db 1023 GlyAlaGluThrPheTyThrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAla 1042  
QY GGCTAGCTGACCGAGCGCGGCGCGCAGAGATCTGTGACCTGACCGAGACCAACCAAG 2097  
Db 1043 GlyTyThrValThrAsnLysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGln 1062  
QY AAGACCGAGCTGCAGGCGCATCCAGTGGCCCTGACGAGCAGCGGCGAGGTGAACATC 2157  
Db 1063 LysThrGluLeuGlnAlaIleTyThrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIle 1082  
QY GTGACCGAGACCGAGTACGCGCTGGCATCATCCAGGCGCCAGCGCCGAGAGGAGGAGC 2217  
Db 1083 ValThrAspSerGlnTyThrAlaLeuGlyIleIleGlnAlaGlnProAspArgSerGluSer 1102  
QY GAGCTGTGACCATCATCATCGACGCTGATCAAGAGGAGAGGTCTACCTGAGCTGG 2277  
Db 1103 GluLeuValSerGlnIleIleGlnLeuIleLysLysGluLysValTyLeuAlaTrp 1122  
QY GTGCGCGCCCAAGAGGCGATCGCGGCGCAACGAGCAGATCCGACAGCTGGTGGAGCAAGGC 2337  
Db 1123 ValProAlaHisLysGlyIleGlyLysGlnGluValAspLysLeuValSerAlaGly 1142  
QY ATCCGCAAGTGTGTTCTCTCGACGCGCATCAT 2370

Db 1143 IleArgLysValLeuPheLeuAspGlyIleAsp 1153  
RESULT 2  
ADP84803  
ID ADP84803 standard; protein; 3183 AA.  
XX ADP84803;  
AC ADP84803;  
XX 23-SEP-2004 (first entry)  
DT  
XX HIV-1 hybrid isolate pNL4-3 protein.  
DE  
XX eliciting; inducing; immune response; HIV; antigen; non-pathogenic;  
KW vaccination; vaccine; HIV-I.  
XX Human immunodeficiency virus 1.  
OS  
XX WO2004056391-A1.  
PN  
XX 08-JUL-2004.  
PD  
XX 19-DEC-2003; 2003WO-AU001705.  
PF  
XX 20-DEC-2002; 2002AU-00953556.  
PR  
XX 17-SEP-2003; 2003AU-00905067.  
XX (UYNE-) UNIV NEW SOUTH WALES.  
PA  
XX Kent SJ, Purcell DF, Boyle DB, Ramsay A, Thomson S, Ramshaw IA;  
PI WPI; 2004-500267/47.  
XX N-PSDB; ADP84804.  
DR  
XX Eliciting or inducing in a mammal an immune response against HIV-1  
PT subtype AE, B or C by administering sequential doses of a recombinant  
PT plasmid and viral vectors containing the nucleic acid molecules encoding  
XX the HIV antigens.  
PS Disclosure; Fig 12; 280pp; English.  
XX The invention relates to a novel method for eliciting or inducing in a  
XX mammal an immune response directed to a virus, preferably Human  
XX immunodeficiency virus (HIV). The method comprises sequentially  
XX administering to the mammal one or more sequential doses of a recombinant  
XX plasmid vector or recombinant viral vector or its derivative, into which  
XX the nucleic acid molecules encoding all, part or a modified form of two  
XX or more antigens of the virus are incorporated, and one or more optimized  
XX CpG motifs, where the antigens have been rendered substantially non-  
XX pathogenic. The invention further comprises: a method for treating or  
XX preventing viral infection in a mammal; a method of vaccinating a mammal  
XX against a viral pathogen; a method of eliciting or inducing, in a mammal,  
XX an immune response directed to HIV; a vaccine capable of inducing an  
XX immune response directed to a virus comprising the recombinant plasmid  
XX vector or recombinant viral vector or its functional derivative; a  
XX nucleic acid construct or its functional derivative comprising the  
XX plasmid vector or recombinant viral vector; a plasmid vector; a  
XX recombinant viral vector; a pharmaceutical composition comprising the  
XX nucleic acid, plasmid vector or recombinant viral vector constructs. The  
XX method is useful in eliciting or inducing in a mammal an immune response  
XX against HIV-1 subtype AE, B or C. This sequence represents the protein of  
XX the HIV-1 hybrid isolate pNL4-3 DNA of the invention.  
SQ Sequence 3183 AA;  
Alignment Scores:  
Pred. No.: 1.94e-213 Length: 3183  
Score: 3782.50 Matches: 711  
Percent Similarity: 87.89% Conservative: 44  
Best Local Similarity: 82.77% Mismatches: 31  
Query Match: 82.53% Indels: 73  
DB: 8 Gaps: 5  
US-09-610-313B-30 (1-2469) x ADP84803 (1-3183)

QY	13	ATGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCTCTGATGAGCGCAGCAAC	69
DB	363	LeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLysGlyAsn	382
QY	70	TTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTCGCGCAAGGAGGCCACATCGCC	129
DB	383	PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla	402
QY	130	CGCAACTCCCGCGCCCCCGCAAGAAGGGCTGCTGGAAAGTCGGCAAGAGGCCACCAG	189
DB	403	LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln	422
QY	190	ATGAAGGACTCGACCGAGCGCCAGGCCAAC	219
DB	423	MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys	442
QY	220	-----TTCTTCCTCCGCGAGAC	234
DB	443	GlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPhePheArgGluAsp	462
QY	235	CTGGCTTTCCTCCAGGGCAAGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGC	294
DB	463	LeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSer	482
QY	294	-----	294
DB	483	ProProGluGluSerPheArgPheGlyGluGluThrThrThrProSerGlnLysGlnGlu	502
QY	295	-----CCACACGCGCGAGCTGCAGGTG	318
DB	503	ProIleAspLysGluLeuTy-ProLeuAlaProThrArgArgGluLeuGlnValTrpGly	522
QY	319	CGCGCGCAACCCCGCGAGCGAGCGCGCGCGAGCGCCAGGGCACCCCTGAAC	372
DB	523	ArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSer	542
QY	373	-----TTCCCCCAGATCACCCCTGTGG	393
DB	543	SerLeuArgSerLeuPheGlySerAspProSerSerGlnPheProGlnIleThrLeuTrp	562
QY	394	CAGCGCCCCCTGGTGAGCATCAAGTGGCGCGCCAGATCAAGGAGGCCCTGTGCACACC	453
DB	563	GlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThr	582
QY	454	GGCGCCGACGACACCGTCTCTGGAGGAGATGACCTCGCCGCGCAAGTGGAGGCCCAAGTG	513
DB	583	GlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysProLysMet	602
QY	514	ATCGCGGCATCGCGGGCTTCATCAAGTGGCCAGTACGACAGATCCTGATCAGATC	573
DB	603	IleGlyGlyIleGlyPheIleLysValArgGlnTyAspGlnIleLeuIleGluIle	622
QY	574	TGCGGCAAGAGGCCATCGGCACCGCTGTGATCGCGCCCAACCCCGGTGAACATCATCGGC	633
DB	623	CysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGly	642
QY	634	CGCAACATGTCACCCAGCTGGCTGCACCTCGAATCTCCCATCAGCCCATCGAGCC	693
DB	643	ArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThr	662
QY	694	GTGCGCGTGAAGCTGAAGCCCGGCATGGACGGCCCAAGGTGAAGCAGTGGCGCCCTGACC	753
DB	663	ValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThr	682
QY	754	GAGGAGAAGATCAAGGCCCTGACCCCATCTCTCGAGGAGATGGAGAGGGCGCAAGATC	813
DB	683	GluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIle	702
QY	814	ACCAAGATCGCGCCGAGAACCCCTACAAACCCCGCTGTTCGCCCATCAAGAAGAAGGAC	873
DB	703	SerLysIleGlyProGluAsnProTyraAsnThrProValPheAlaIleLysLysLysAsp	722

QY	874	AGCACCAAGTGGCGCAAGCTGGTGGACTTCCGGAGCTGAACAAGCGCACCCAGGACTTC	933
DB	723	SerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPhe	742
QY	934	TGGGAGGTGCAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAAGACGAGCGTGACC	993
DB	743	TrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysSerValThr	762
QY	994	GTGCTGACGTGGCGCAGCGCTACTTACGCGTCCCTTGGACGAGGACTTCCGCAAGTAC	1053
DB	763	ValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheAsgLysTy	782
QY	1054	ACCGCTTACCATCCCGACATCAACAACGAGACCCCGGCATCGCTACCACTACAAC	1113
DB	783	ThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGlnTy	802
QY	1114	GTGCTGCCCGAGGGCTGGAAGGCGAGCCCGACGATCTTCCAGAGCAGCATGACCAAGATC	1173
DB	803	ValLeuProGlnTyTrpLysGlySerProAlaIlePheGlnCysSerMetThrLysIle	822
QY	1174	CTGGAGCGCTTCCGCGCCCGCAACCCCGAGATCGTGATCTTACCAGTACATGAGCAGCTG	1233
DB	823	LeuGluProPheArgLysGlnAsnProAspIleValIleTyGlnTyMetAspAspLeu	842
QY	1234	TACGTGGCGACGACTGGAGATCGGCCACGACCGCGCCCAAGATCGAGGAGCTGCGCAAG	1293
DB	843	TyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGln	862
QY	1294	CACCTGTGCGCTGGGGCTTACCACCCCGACAAAGAACGACACGAGGAGCGCCCTTC	1353
DB	863	HisLeuLeuArgTrpGlyPheThrThrProAspLysHisGlnLysGluProProPhe	882
QY	1354	CTGTGATGGCTACGAGCTGCACCCCGCAAGTGGACCGCTGACGCCATCGAGCTGCC	1413
DB	883	LeuTrpMetGlyTyGluLeuHisProAspLysTrpThrValGlnProIleValLeuPro	902
QY	1414	GAGAAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGTGGCGAAGCTGAAGCTGG	1473
DB	903	GluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAla	922
QY	1474	AGCCAGATCTACCCCGCATCAGGTGCGCGACGCTGTGCAAGCTGTGCGCGGCCAAG	1533
DB	923	SerGlnIleTyAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLys	942
QY	1534	GCCCTGACGACATGTGTGCCCTTGACCGAGGAGCGCAGCTGGAGCTGGCGCGAGACCG	1593
DB	943	AlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArg	962
QY	1594	GAGATCTCGCGCAGCCGCTGGACCGCGTGTACTACGACCCGACGAAGGACCTGTGGCC	1653
DB	963	GluIleLeuLysGluProValHisGlyValTyTyAspProSerLysAspLeuIleAla	982
QY	1654	GAGATCCAGAAGGCGCCAGCAGTGCACCTTACCAGATCTTACCAGGAGCGCTTCAAG	1713
DB	983	GluIleGlnLysGlnGlyGlnTyTrpThrTyGlnIleTyGlnGluProPheLys	1002
QY	1714	AACCTGAAGACCGCAAGTAGTCGCAAGATCGCACCGCCACCAACGACGCTGAAGCAG	1773
DB	1003	AsnLeuLysThrGlyLysTyAlaArgMetCysGlyAlaHisThrAsnAspValLysGln	1022
QY	1774	CTGACCGAGGCGGTGCAAGAAGATCGCCATCGGAGAGCATCGTGAGCCAGACCCCC	1833
DB	1023	LeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPro	1042
QY	1834	AAGTTCGCGCTGCCCATCAGAGGAGACCTGGGAGACCTGTGGACCGCACTACTGCGAG	1893
DB	1043	LysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGlnTyTrpGln	1062
QY	1894	GCCACCTGGATCCCGAGTGGGAGTTCGTGAAACACCCCGCTGGTGAAGCTGTGGTAC	1953
DB	1063	AlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTy	1082
QY	1954	CAGCTGGAGAGGAGGCCCATCATCGCGCGCAGAGACTTCTACGTGAGCGCGCGCCCAAC	2013

```

Db      1083  GinLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAlaAsn 1102
Qy      2014  CGCGAGCAAGATGGCAAGCCCGCTACCTGACCGACCGCGCGCGCGAGAAATCGTG 2073
Db      1103  ArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValVal 1122
Qy      2074  AGCTGTACCGAGACCAACCAAGACAGACAGCCAGCTGACGGCCATCCAGCTGGCCCTGCAG 2133
Db      1123  ProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142
Qy      2134  GACAGCGCGCAGCGAGTGAACATCGTGACACAGCCAGTACGCGCTGGGATCATCCAG 2193
Db      1143  AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGln 1162
Qy      2194  GCCAGCGCCGACAAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAG 2253
Db      1163  AlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuLys 1182
Qy      2254  AAGGAGAAGGTGTACCTGAGCTGGGTGCGCCGCCCCACAAAGGCATCGCGGCAAGCAG 2313
Db      1183  LysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGln 1202
Qy      2314  ATCGACAGCTGGTGAGCAAGGGATCGCAAGGTGCTGTTCTTGGAGCGGCATCGAT 2370
Db      1203  ValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1221

RESULT 3
ADN36406
ID      ADN36406 standard; protein; 1457 AA.
AC      ADN36406;
XX      15-JUL-2004 (first entry)
XX      Human protein for anti-HIV vaccine.
XX      anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX      Homo sapiens.
XX      WO2004035006-A2.
XX      29-APR-2004.
XX      17-OCT-2003; 2003WO-US033112.
XX      18-OCT-2002; 2002US-0419465P.
XX      (AARO-) AARON DIAMOND AIDS RES CENT.
XX      Huang Y, Ho DD, Chen Z;
XX      WPI; 2004-348328/32.
XX      N-PSDB; ADN36405.
XX      Nucleic acid vector comprising at least one HIV sequence operably linked
XX      to a promoter and encoding a protein that does not assemble into viral
XX      particles, useful in immunizing a subject against HIV infection.
XX      Disclosure; SEQ ID NO 20; 166pp; English.
XX      The invention relates to a nucleic acid vector comprising at least one
XX      HIV sequence operably linked to a promoter and encoding a protein that
XX      does not assemble into viral particles. The nucleic acid vector is useful
XX      in immunizing a subject against HIV infection. This sequence corresponds
XX      to a peptide used in the invention.
XX      Sequence 1457 AA;

Alignment Scores:
Pred. No.: 4.08e-211 Length: 1457
Score: 3741.50 Matches: 709

```

```

Percent Similarity: 92.24% Conservative: 28
Best Local Similarity: 88.74% Mismatches: 44
Query Match: 81.64% Indels: 19
DB: 8 Gaps: 3
US-09-610-313B-30 (1-2469) x ADN36406 (1-1457)

Qy      /13  ATGCGCCAGCCATGAGCGAGGCCACCGCCGACCATCTGATGCGCGCAGCAACTTC 72
Db      383  LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnPhe 401
Qy      73   AAGGGCCCCAAGCCATCATCAAGTGTTCACCTGCGCGAGAGAGAGGCCACATCGCCCGC 132
Db      402  LysGlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 421
Qy      133  AACTGCGCGCCGCCCAAGAGAGGCTGTGGAGTGCAGGAAAGAGGGGCCACACAGATG 192
Db      422  AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 441
Qy      193  AAGCACTGCACCGCCAGCCAGCCCAACTTCTTCGCGAGGACCTGGGCTTCCCCAGGGC 252
Db      442  LysAspCysThrGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProProHisLysG 461
Qy      253  AAGCCCGCGAGTTCGCCAGCAGAAACCGCCCAACAGCCACACAGCCCGCGAGCTG 312
Db      461  YArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPh 481
Qy      313  CAGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGAGCCCGAGGGCACCTGAAAC 372
Db      481  eGlyPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuTh 501
Qy      373  TTC-----CCCCAGATCACCCTGTGGCA 395
Db      501  rSerLeuLysSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpG 521
Qy      396  GCGCCCTCGTGAGCATCAAGTGGCGCGCCAGATCAAGAGAGCCCTGCTGGACACCGG 455
Db      521  nArgProLeuValSerIleArgValGlyGlnIleLysGluAlaLeuLeu----- 538
Qy      456  CGCGCAGCACACCGTGTGGAGGAGATGAGCTCCCGCAAGTGGAAAGCCCAAGATGAT 515
Db      539  ----AspAspThrValLeuGluValAsnLeuProGlyLysTrpLysProLysMetIl 557
Qy      516  CGCGCGCATCGCGGCTTCATCAAGTGGCGCGAGTACGACAGATCTGTAGATCGATCTG 575
Db      557  eGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleProIleGluIleCy 577
Qy      576  CGGCAAGAGGCCATCGGCACCGTGTGATCGGCGCCCGCCAGCCCGCTGACATCATCGCG 635
Db      577  sGlyLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597
Qy      636  CAACATGCTGACCCAGCTGGCTGCACCTCAACTTCCCATCAGCCCGCATCGAGACCGT 695
Db      597  gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrIl 617
Qy      696  GCCCGTCAAGCTGAAGCCCGCATGAGCGCGCCCAAGTGAAGCAGTGGCCCTGACCGA 755
Db      617  eProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGl 637
Qy      756  GGAGAAGATCAAGCCCTGACCGCCATCTGGAGGAGATGCGAGAGGAGGCGGCAAGTAC 815
Db      637  uGluLysIleLysAlaLeuThrAlaIleCysAspGluMetGluLysGluGlyLysIleTh 657
Qy      816  CAAGATCGCGCCCGAGAACCCCTACAACACCCCGCTGTTCGCCATCAAGAAAGAGGACAG 875
Db      657  rLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSe 677
Qy      876  CACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCGCAACCCAGGACTTCTG 935
Db      677  rThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTr 697
Qy      936  GGAGGTGAGCTGGGCATCTCCCGACCCCGCGCTGAAGAGAGAGAGCGGTGACCGT 995

```



697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysSerValThrVa 717  
QY 996 GCTGGAGCTGGCGACCCCTACTTACGCTGCCCTGGACGAGACTTCCGCAAGTACAC 1055  
Db 717 lLeuAspValGlyAspAlaTy rPheSerValProLeuTy rGluAspPheArgLysTy rTh 737  
QY 1056 CGCCTTCCACCATCCCCAGCATCAACACAGAGACCCCGGCATCCGCTTACCAAGTACAACT 1115  
Db 737 rAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTy rGlnTy rAsnVa 757  
QY 1116 GCTGCCCCAGGCTGGAGGGAGCCCCAGCATCTTCCAGAGCATGACCAAGATCCT 1175  
Db 757 lLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777  
QY 1176 GGAGCCCTCCCGCGCCGCAACCCCGAGATCGTGATCTACAGTACATGAGCAGCACCTGTA 1235  
Db 777 uGluProPheArgAlaGlnAsnProGluIleValIleTy rGlnTy rGlyAspAspLeuTy 797  
QY 1236 CGTGGGAGGAGCTGGAGATCGGCCAGCACCGCGCCCAAGATCGAGAGCTGGCGCAAGCA 1295  
Db 797 rValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHI 817  
QY 1296 CTGTGCTGGCTGGGCTTACACCCCGGCAAGAGACCAAGAGAGCCCGCTTCCT 1355  
Db 817 sLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLe 837  
QY 1356 GTCGATGGGTACGAGTGCACCCCGACAGTGCAGCGTGCAGCCATCGAGCTGCCCGA 1415  
Db 837 uTrpMetGlyTy rGluLeuHisProAspLysTy rThrValGlnProIleGlnLeuProGI 857  
QY 1416 GAAGGAGAGCTGGACCCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCCAG 1475  
Db 857 uLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877  
QY 1476 CCGATCTACCCCGGCATCAAGGTGGCCAGCTGTGTGCAAGCTCTGCGCGGCCAGGC 1535  
Db 877 rGlnIleTy rProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897  
QY 1536 CCGTACCGCATCGTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCGA 1595  
Db 897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGI 917  
QY 1596 GATCTCGCGAGCCCGTGCACCGCGTGTACTACGACCCCGACGAGCAACCTGGTGGCCGA 1655  
Db 917 uIleLeuLysGluProValHisGlyAlaTy rAspProSerLysAspLeuIleAlaGI 937  
QY 1656 GATCCAGAACGAGGCCACACAGCTGAGCTTACAGATCTACAGAGCCCTTCAAGAA 1715  
Db 937 uIleGlnLysGlnGlyGlnAspGlnTrpThrTy rGlnIleTy rGlnGluProPheLysAs 957  
QY 1716 CCGTGAAGCCGCAAGTACCGCAAGATGCGACCGCCACACCAACGAGCTGAAGCAGCT 1775  
Db 957 nLeuLysThrGlyLysTy rAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977  
QY 1776 GACGAGGCCCTGCAGAAATCCGATCGATCGAGAGATCGTATCTGGGGCAAGACCCCCAA 1835  
Db 977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTrpGlyLysIleProLy 997  
QY 1836 GTTCCGCTCCCATCCAGAGAGAGACTGGGAGACTGTGTGACCGACTACTGGCAGGC 1895  
Db 997 sPheArgLeuProIleProLysGluThrTrpGluThrArgTrpThrAlaTy rTrpGlnAl 1017  
QY 1896 CACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGTACCA 1955  
Db 1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTy rGI 1037  
QY 1956 GCTGGAGAGAGCCCATCATCGCGCCGAGACTTCTACGTGGAACCGCCCGCCCAACCG 2015  
Db 1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTy rValAspGlyAlaAlaAsnAr 1057  
QY 2016 CGAGACCAAGATCGGCAGCCCGCTACGTGACCGACCGCGCGCGGAGAGATCGTGAG 2075  
Db 1057 gGluThrLysMetGlyLysAlaGlyTy rValThrAspArgGlyArgGlnLysIleValSe 1077

QY 2076 CTTGACCGAGACCAACCAAGAGACCGAGCTGCGAGCCATCCAGCTGCGCCCTGCAGGA 2135  
Db 1077 rLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleCysLeuAlaLeuGlnAs 1097  
QY 2136 CAGCGCAGCAGGAGTGAACATCGTGACCGACAGCAGTACGCGCTGGGATCATCCAGGC 2195  
Db 1097 pSerGlySerGluValAsnIleValThrAspSerGlnTy rAlaLeuGlyIleIleGlnAl 1117  
QY 2196 CCAGCCCGACAGAGCAGAGCGAGCTGCTGAACAGATCATCGAGCAGCTGATCAAGAA 2255  
Db 1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuLysLy 1137  
QY 2256 GGAGAAGGTGTACTCGAGCTGGTGGTCCGCCACCAAGGGCATCGCGGCAACCGAGCAGAT 2315  
Db 1137 sGluArgValTy rLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVa 1157  
QY 2316 CGAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCCTGGACGGCATCGAT 2370  
Db 1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175

RESULT 4  
AAB69289  
ID AAB69289 standard; protein; 1003 AA.  
XX AAB69289;  
XX  
DT 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX  
DE HIV-1 non-subtype B clone 94IN476-104 pol protein.  
XX  
XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
KW vif; vpr; tat; rev; nef; vaccine.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO200026416-A1.  
XX  
PD 11-MAY-2000.  
XX  
PF 25-OCT-1999; 99WO-US024837.  
XX  
PR 02-NOV-1998; 98US-00184418.  
XX  
XX (UABR-) UAB RES FOUND.  
XX  
XX Hahn BH, Shaw GM, Gao F;  
XX WPI; 2000-365651/31.  
XX  
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.  
XX  
XX  
PS Claim 41; Fig 15; 131pp; English.  
XX  
XX The present invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 1003 AA;

Alignment Scores:  
Pred. No.: 2,96e-206 Length: 1003  
Score: 3658.00 Matches: 681  
Percent Similarity: 98.61% Conservative: 26  
Best Local Similarity: 94.98% Mismatches: 10

Query Match:	79.82%	Indels:	0
DB:	3	Gaps:	0
US-09-610-313B-30 (1-2469) x AAB69289 (1-1003)			
QY	220	TTCTTCCGCGAGGACCTGGCTTCCCTCCAGGCGAAGCCGCGAGTTCCTCCAGCGAGCAG	279
DB	1	PhePheArgGluAsnLeuAlaPheProGlnGlnGluAlaArgGluPheProSerLysGln	20
QY	280	AACCGCGCCAAAGCCCAACGCGCGAGCTGCGAGTGCAGGCGCGCGCAACAACCCCGCGAGC	339
DB	21	AlaArgAlaAsnSerProThrSerArgGluLeuGlnValGlnGlnGlyAspAsnProArgSer	40
QY	340	GAGCGCGCGCGAGCGCGAGCGCACTGAACTTCCCGCGAGATCACCTGTGGCAGCGC	399
DB	41	GluAlaGlyValGluArgGlnGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg	60
QY	400	CCCTGTGTGACATCAAGGTGGCGCGCAGATCAGAGGCGCCCTCTGCGACACCGCGCC	459
DB	61	ProLeuValSerIleLysValGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla	80
QY	460	GACGACACCGTGTGCGAGGAGATGAGCTGCGCGCAAGTGGAAAGCCCAAGATGATCGC	519
DB	81	AspAspThrValLeuGluGluLeuAlaLeuProGlyArgTrpLysProLysMetIleGly	100
QY	520	GGCATCGCGCGCTTCATCAAGGTGGCGCGAGTACGACCAAGATCTCTGATCGAGATCTGCGC	579
DB	101	GlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGly	120
QY	580	AAGAAGGCATCGGCACCGCTGATCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGC	639
DB	121	LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsp	140
QY	640	ATGTGTGACCGAGCTGGGCTGCACCTGAATCTCCCACTGCGCGCGCGCGCGCGCGCGCGC	699
DB	141	MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro	160
QY	700	GTGAAGCTGAAGCCCGGCGATCGAGCGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGAG	759
DB	161	VallLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu	180
QY	760	AAGATCAAGGCGCTGACCGCTATCGCGAGAGATGGAGAGGAGGCGCGAGATCACCAAG	819
DB	181	LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys	200
QY	820	ATCGCGCGCGAGAACCCCTACAACACCGCGTGTTCGCGCATCAAGAAGAGAGCAGACAC	879
DB	201	IleGlyProGluAsnProTyAsnThrProValPheAlaIleLysArgLysAspSerThr	220
QY	880	AAGTGGCGCAAGCTGTGGACTTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAG	939
DB	221	LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu	240
QY	940	GTGAGCTGGCGATCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	999
DB	241	ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu	260
QY	1000	GAGTGGCGGACGCTACTTCAAGCGTGCCTGACGAGGACTTCCGCAAGTACACCGCC	1059
DB	261	AspValGlyAspAlaTyPheSerValProLeuAspGluGlyPheGlyLysTyThrAla	280
QY	1060	TTCAACCTCCCGAGCATCAACAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	1119
DB	281	PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGlnTrpAsnValLeu	300
QY	1120	CCCAGGCGTGAAGGC	1179
DB	301	ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu	320
QY	1180	CCCTTCCGC	1239
DB	321	ProPheArgAlaArgAsnProLysIleValIleTyGlnTrpMetAspLeuTyVal	340

QY	1240	GGCAGGACCTGGAGATCGCGCAGCACCGCGCCCAAGATCGAGGAGCTGGCGAAGCACCTG	1299
DB	341	GlySerAspLeuGluIleGlyHisArgAlaLysIleGluGluLeuArgAlaHisLeu	360
QY	1300	CTGCGCTGGGCTTCACCAACCCCGCAAGAACACCAAGAGGAGGCCCTTCTCTGTGG	1359
DB	361	LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp	380
QY	1360	ATGGGCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAAG	1419
DB	381	MetGlyTyGluLeuHisProAspLysTrpThrValGlnProIleLysLeuProGluLys	400
QY	1420	GAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGCAAGCTGAACCTGGCGCGCAG	1479
DB	401	AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln	420
QY	1480	ATCTACCCCGCATCAAGGTGGCGCGCTGTGCAAGCTGTGCGCGCGCGCAAGCCCTG	1539
DB	421	IleTyProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu	440
QY	1540	ACCAGCATCTGTCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAACCGCGAGATC	1599
DB	441	ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIle	460
QY	1600	CTGCGCGAGCCCGTGCAGCGGTGTACTACACCCCGCAAGAGACCTGGTGGCGAGATC	1659
DB	461	LeuLysGluProValHisGlyValTyTrpAspProSerLysAspLeuIleAlaGluIle	480
QY	1660	CAGAAGCAGGCGCACGACGAGTGCACCTACAGATCTACAGAGGAGCCCTCAAGAACCTG	1719
DB	481	GlnLysGlnGlyHisAspGlnTrpThrTyGlnIleTyGlnGluProPheLysAsnLeu	500
QY	1720	AAGACCGCGCAAGTACGCGCAAGATCGCACCGCCCGCACCAACGACGCTGAAGCAGCTGACC	1779
DB	501	LysThrGlyLysTyAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr	520
QY	1780	GAGCGCTGCAGAAGATCGCCATCGAGAGCATCTGTATCTGGGGCAAGACCCCGAGTTC	1839
DB	521	GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp**LysThrProLysPhe	540
QY	1840	CGCTGCGCCATCCAGAAGGAGAGCTGGGAGACCTGGTGGACCGACTACTGCGACGCCACC	1899
DB	541	ArgLeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyTrpGlnAlaThr	560
QY	1900	TGGATCCCGCAGTGGGAGTTCGTGAACACCCCGCCCTCTGTGTGAAGCTGTGTACAGCTG	1959
DB	561	TrpIleProAspTrpGluPheValAsnThrProProLeuValLysLeuTyTrpGlnLeu	580
QY	1960	GAGAAGGAGCCATCATCTCGCGCGCGAGACCTTCTACGTGAGCGGCGCGCGCAACCGCGAG	2019
DB	581	GluLysGluProIleValGlyAlaGluThrPheTyValAspGlyAlaAlaAsnArgGlu	600
QY	2020	ACCAAGATCGCGAGCGCGCTAGTACCGACCGCGCGCGCGCGAGAGATCGTGGAGCTG	2079
DB	601	ThrLysValGlyLysAlaGlyTyThrAspArgGlyArgGlnLysIleValSerLeu	620
QY	2080	ACCAGAGCACCAACCAAGAAGACCGAGCTGCAGGCCATCCAGCTGGCGCCCTCGACGACAGC	2139
DB	621	ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer	640
QY	2140	GGCAGCGAGTGAACATCTGTGACCGCAGCAGCTACGCCCTGGGCATCATCCAGGCCCGAG	2199
DB	641	GlyThrGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIleIleGlnAlaGln	660
QY	2200	CCGAGAGGCGAGGAGCTGTGTAACAGATCATCGAGCTGAGCTGATCAAGAGGAG	2259
DB	661	ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleAsnLysGlu	680
QY	2260	AAGGTGTACCTGAGCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	2319
DB	681	ArgValTyLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnGluValAsp	700
QY	2320	AAGTGTGAGCAAGGCGCATCCCGAAGGTGCTGTCTCTGCGCGCGCATCGAT	2370

```
Db 701 ArgLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717
:::|||||
RESULT 5
AAB69286
ID AAB69286 standard; protein; 999 AA.
XX
XX AAB69286;
XX AC
XX 12-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
XX
XX HIV-1 non-subtype B clone 962W651-8 pol protein.
DE
DE HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;
KW vif; vpr; tat; rev; nef; vaccine.
XX
XX Human immunodeficiency virus 1.
OS
XX
XX WO200026416-A1.
PN
XX
XX 11-MAY-2000.
PD
XX
XX 25-OCT-1999; 99WO-US024837.
PF
XX
XX 02-NOV-1998; 98US-00184418.
PR
XX
XX (UABR-) UAB RES FOUND.
PA
XX
XX Hahn BH, Shaw GM, Gao F;
PI
XX
XX WPI; 2000-365651/31.
DR
XX
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT type 1 useful for detecting and treating AIDS comprises a specific
PT nucleotide sequence.
XX
XX Claim 41; Fig 15; 13lpp; English.
PS
XX
XX The present in invention provides the protein and coding sequences for a
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC rev and nef proteins. These can be used to detect the presence of HIV-1
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC These antibodies can be used in vaccines to prevent and treat HIV
CC infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
XX Sequence 999 AA;
SQ
Alignment Scores:
Pred. No.: 9,97e-206 Length: 999
Score: 3649.00 Matches: 24
Percent Similarity: 98.33% Conservative: 24
Best Local Similarity: 94.98% Mismatches: 12
Query Match: 79.62% Indels: 0
DB: 3 Gaps: 0
US-09-610-313B-30 (1-2469) x AAB69286 (1-999)
QY 220 TTCTTCGCGAGGACCTGGGCTTCCCGGCAAGCGCGGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20
QY 280 AACCGCGCCAAACGCCACACAGCGGAGCTGCGGCGGCGGAGAACCCCGGAGC 339
Db 21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSer 40
QY 340 GAGCGCGCGCCGAGCGCGAGGCGACCCCTGAATTCCTCCCGAGATCACCCTGTGCGCGC 399
Db 41 GluAlaGlyValGluArgGlnGlySerLeuAsnPheProGlnIleThrLeuTrpGlnArg 60
QY 400 CCCTGTGTGATCATCAAGGTGGCGCGCCAGATCAAGGAGGCCCTGTGTGGACACCGCGGCC 459
Db 61 ProLeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuAspThrGlyAla 80
QY 460 GACGACACCGTGTCTGAGGAGATGACCTCCCGGCAAGTGGAGGCCAAGATGATCGGC 519
Db 81 GlyAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGly 100
QY 520 GGCATCGGGGCTTCATCAAGGTGGCGGACGACGACGATCCTGATCAGATCTCGGC 579
Db 101 GlyIleGlyPheIleGluValArgGlnTyrAspGlnIleProMetGluIleCysGly 120
QY 580 AAGAAGCCATCGGCACCGTGTGATCGGCCCCACCCCGTGAACATCATCTCGCGCGCAAC 639
Db 121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 140
QY 640 ATGCTGACCCAGCTGGGCTGCACTTCCGAGTCCCATCAGCCCATCAGAGCCGTCGCC 699
Db 141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160
QY 700 GTGAAGCTGAAGCCCGCATGACGCGCCCAAGGTGAAGCAGTGGGCCCTGACCGGAGG 759
Db 161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180
QY 760 AAGATCAAGGCCCTGACCCGCTCTCGGAGGAGATGGAGAGGAGGCGCAAGATCACCAAG 819
Db 181 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 200
QY 820 ATCGGCCCGGAGAACCCCTACACACCCCGCTGTCGCCATCAAGAGGAGGACGACGACC 879
Db 201 IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThr 220
QY 880 AAGTGGCGCAAGCTGGTGACTTCCCGGAGCTGAACAAGCGCACCCGAGGACTTCTGGGAG 939
Db 221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240
QY 940 GTGACGCTGGGCATCCCCACCCCGCGCTGGAAGAAGAAGAGCGTACCGTCTGCTG 999
Db 241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu 260
QY 1000 GAGTGGCGGCGCCTTACTTCCAGCTGCGCCCTGAGGAGGACTTCCGCAAGTACACGCC 1059
Db 261 AspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysIleThrAla 280
QY 1060 TTCACCATCCCCAGCATCAACAACAGAGACCCCGGATCCGCTACCAAGTACAAACGCTG 1119
Db 281 PheThrIleProSerThrAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 300
QY 1120 CCCAGGCTGGAAGGCGCGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAG 1179
Db 301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320
QY 1180 CCCTTCGCGCGCGCAACCCGAGATCGTGATCTACCGATACATGACGACGACCTGTAGTG 1239
Db 321 ProPheArgAlaGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTrpVal 340
QY 1240 GGCAGCGACCTGAGATCGCGCCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGACACCTG 1299
Db 341 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 360
QY 1300 CTGCGTGGGGCTTACCAACCCCGCAAGAAGCACACAGAGGAGGCCCTTCTGTGTGG 1359
Db 361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380
QY 1360 ATGGGTACGAGCTGACCCCGCAAGTGGACCGTGCAGCCCATCAGCTGCCCGCGAGAG 1419
Db 381 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnAlaGluLys 400
QY 1420 GAGAGCTGACCGCTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGGCGAGCCAG 1479
Db 401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420
QY 1480 ATCTACCCCGGCATCAAGGTGGCGCCAGCTGTGCAAGCTGCTGCGCGCGCCCAAGGCCCTG 1539
```

Db 421 IleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440  
 Qy 1540 ACCGACATCGTGGCCCTGACCGAGAGGCCAGCTGGAGCTGGCCGAGAACCGCGAGATC 1599  
 Db 441 ThrAspIleValProLeuThrGluGluAlaGluLeuLeuAlaGluAsnLysGluIle 460  
 Qy 1600 CTGCGCGAGCCGCTGACGGCTGTACTACGACCCAGCAGGACCTGGTGGCGAGATC 1659  
 Db 461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480  
 Qy 1660 CAGAAGCAGGCGCCACGACCTGACCTACAGATCTACACGAGGCCCTTCAAGAACCTG 1719  
 Db 481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500  
 Qy 1720 AAGACCGCAAGTACGCAAGATGCGCAACCGCCACACACGACCTGGAAGCAGCTGACC 1779  
 Db 501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520  
 Qy 1780 GAGCGCTGCAGAGATCGCCATCGGAGAGCATCTGATCTGGGCGAAGACCCCAAGTTC 1839  
 Db 521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTrpGlyLysIleProLysPhe 540  
 Qy 1840 CGCTGCGCCATCCAGAGCAGACCTGGGAGACCTGGTGACCGACTACTGGCAGCCACCC 1899  
 Db 541 ArgLeuProLysGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAlaThr 560  
 Qy 1900 TGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTG 1959  
 Db 561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTrpTyrGlnLeu 580  
 Qy 1960 GAGAAGGAGCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAG 2019  
 Db 581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600  
 Qy 2020 ACCAAGATCGCAGGCGCGCTAGCTGACCGACCGCGCGCGCAGAGATCGTGAGCTG 2079  
 Db 601 ThrLysLeuGlyLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620  
 Qy 2080 ACCGAGACCCACCAACGAGACCGAGCTGCAGGCCCATCCAGCTGGCGCTGCAGGACAGC 2139  
 Db 621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640  
 Qy 2140 GGCAGCAGGTGAATCATGTCACGACGACGACGACGCGCGCGCGCATCATCCAGGCCGAG 2199  
 Db 641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660  
 Qy 2200 CCGCACAAGACGAGAGCGAGCTGTGAACACGAGATCATCGAGCAGCTGATCAAGAAGGAG 2259  
 Db 661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGlu 680  
 Qy 2260 AAGTGTACCTGAGCTGGGTGCCCGCCACACAAGGGCATCGCGCGCAACGAGCAGATCGAC 2319  
 Db 681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAsp 700  
 Qy 2320 AAGCTGTGACCAAGGGATCCGCAAGGTGCTGTCTCTGGACGGCATCGAT 2370  
 Db 701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 6  
 AAB69287  
 ID AAB69287 standard; protein; 1005 AA.  
 XX  
 AC AAB69287;  
 XX  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 XX HIV-1 non-subtype B clone 962M751-3 pol protein.  
 XX  
 XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 XX  
 OS Human immunodeficiency virus 1.

XX WO200026416-A1.  
 PN 11-MAY-2000.  
 PD 25-OCT-1999; 99WO-US024837.  
 PF 02-NOV-1998; 98US-00184418.  
 PR (UABR-) UAB RES FOUND.  
 PA Hahn BH, Shaw GM, Gao F;  
 PI WPI; 2000-365651/31.  
 DR Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 XX type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.  
 PT Claim 41; Fig 15; 13lpp; English.  
 PS The present invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1005 AA;  
 Alignment Scores:  
 Pred. No.: 1.96e-205 Length: 1005  
 Score: 3644.00 Matches: 685  
 Percent Similarity: 97.23% Conservative: 18  
 Best Local Similarity: 94.74% Mismatches: 14  
 Query Match: 79.51% Indels: 6  
 DB: 3 Gaps: 1  
 US-09-610-313B-30 (1-2469) x AAB69287 (1-1005)  
 Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyGluLeuProSerGluGln 20  
 Qy 280 AACCGCGCC-----AACAGCCCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 321  
 Db 21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40  
 Qy 322 GCGCACAACCCCGCAGCG 381  
 Db 41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60  
 Qy 382 ATCACTCTGTGGCAGCC 441  
 Db 61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIleLysGluAla 80  
 Qy 442 CTGCTGGACACCG 501  
 Db 81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluIleAsnLeuProGlyLysTrp 100  
 Qy 502 AAGCCCAAGATGATCG 561  
 Db 101 LysProLysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIle 120  
 Qy 562 CTGATCGAGATCTGCG 621  
 Db 121 LeuIleGluIleCysGlyLysAlaIleGlyThrValLeuValGlyProThrProVal 140  
 Qy 622 AACATCATCG 681  
 Db 141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160

QY 682 CCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGCATGGACCGGCCCAAGAGTGAAGCAG 741  
 Db 161 ProIleGluThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGln 180  
 QY 742 TGGCCCTGACCGAGGAGAAAGTCAAGCCCTGACCGCCATCTGCGAGGAGATGGAGAAG 801  
 Db 181 TriProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLys 200  
 QY 802 GAGGGCAAGTACCAAGATCGGCGCCCGAGAACCCCTACACACCCCGTGTGCGCATC 861  
 Db 201 GluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIle 220  
 QY 862 AAGAAGAAGACAGACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTTGAACAAGCGC 921  
 Db 221 LysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArg 240  
 QY 922 ACCGAGGACTTCTGGAGGTGCGAGCTGGGATCCCCATCCCCCGCCGCTTGAAGAAGAAG 981  
 Db 241 ThrGlnAspPheTrpGluValGlnLeuLysIleProHisProAlaGlyLeuLysLysLys 260  
 QY 982 AAGAGGTGACCGTGTGAGCGTGGGCGAGCGCTACTTCAAGCTGGCCCTGGAGGAGAC 1041  
 Db 261 LysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGly 280  
 QY 1042 TTCGCGAAGTACACCGCTTCCACATCCCGCATCAACAACAGAGACCCCGCCGATCCGC 1101  
 Db 281 PheArgLysIleThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyLysArg 300  
 QY 1102 TACCAGTACAACTGTGTCGCCCGGAGGCTGGAAGGGAGCGCCAGCATCTTCCAGAGCAGC 1161  
 Db 301 TyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSer 320  
 QY 1162 ATGACCAAGATCTGTGAGCCCTTCCGCGCCGCGCAACCCCGAGATCTGTACTACCACTAC 1221  
 Db 321 MetIleLysIleLeuGluProPheArgThrGlnAsnProGluIleValIleTyrGlnTyr 340  
 QY 1222 ATGAGCAGCTGTACGTGGGCGAGCGACTGGAGATCGGCAGCACCGCGCAAGATCTGAG 1281  
 Db 341 MetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlu 360  
 QY 1282 GAGCTGCGCAAGCACTGTGCTGCGCTGGGGTTCCACACCCCGCAAGAAGACCAAGAAG 1341  
 Db 361 GluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLys 380  
 QY 1342 GAGCCCGCTTCTGTGGATGGGTACGAGCTGACCGCCCGCACCAAGTGGACCGTGCAGCCC 1401  
 Db 381 GluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPro 400  
 QY 1402 ATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAAG 1461  
 Db 401 IleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLys 420  
 QY 1462 CTGAACCTGGGCGCAGCATCTACCCCGCATCAAGGTGCGCGAGCTGTCAAGCTGCTG 1521  
 Db 421 LeuAsnTrp\*\*SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeu 440  
 QY 1522 CGGCGCCCAAGCCCTGACGACATCGTCCCTGACCGAGGCGCGAGCTGGAGCTG 1581  
 Db 441 ArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGluLeuGluLeu 460  
 QY 1582 GCCGAGAACCGGAGATCTTGGCGGAGCGCGTGCACCGCTGTACTACGACCCCGCAGCAAG 1641  
 Db 461 AlaGluSerArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLys 480  
 QY 1642 GACCTGTGGCGGAGATCCAGAAGACGAGCGGCACGACAGTGGACCTACAGATCTACCAAG 1701  
 Db 481 AspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnValTyrGln 500  
 QY 1702 GAGCCCTTCAAGACCTGAAGACCGGCAAGTACCGCAAGATGCGCACCGCCCAACCAAC 1761  
 Db 501 GluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsn 520

QY 1762 GACGTGAACAGCAGCTGATCCAGGCGCGTGCAGAGATGCCATGGAGAGCATCGTGAATCTGG 1821  
 Db 521 AspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTyr 540  
 QY 1822 GGCAGACCCCAAGTTCCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACC 1881  
 Db 541 GlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThr 560  
 QY 1882 GACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTGGTG 1941  
 Db 561 AspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVal 580  
 QY 1942 AAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAC 2001  
 Db 581 LysLeuTyrTrpGlnLeuGluLysGluProIleAlaGlyAlaGluThrTyrTyrValAsp 600  
 QY 2002 GGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGGCCGG 2061  
 Db 601 GlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArg 620  
 QY 2062 CAGAAGATCGTGAGCTGACCGAGACCCCAACCAAGAGACCCAGCTGAGCGGCATCCAG 2121  
 Db 621 GlnLysIleValThrLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGln 640  
 QY 2122 CTGGCCCTCAGGACAGCGGACGAGGTGAACATCGTGAACCGACGACGAGTACGCCCTG 2181  
 Db 641 LeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeu 660  
 QY 2182 GGCATCATCCAGCCCGCGACGAGGAGCGAGCTGGTGAACAGATCATCGAG 2241  
 Db 661 GlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGlu 680  
 QY 2242 CAGCTCATCAAGAAGAGAGAGTGTACTGAGCTGGGTGCCCGCCACAGGCGCATCGGC 2301  
 Db 681 GlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGly 700  
 QY 2302 GCGAAGCAGCAGATCAAGAGCTGGTGAAGGAGATCCGCAAGGTGCTGTCTCTGGAC 2361  
 Db 701 GlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAsp 720  
 QY 2362 GGCATCGAT 2370  
 Db 721 GlyIleAsp 723  
 RESULT 7  
 AAE37601  
 ID AAE37601 standard; protein; 998 AA.  
 AC AAE37601;  
 XX  
 DT 23-OCT-2003 (revised)  
 DT 27-AUG-2003 (first entry)  
 XX  
 DE HIV-1 subtype C isolate Dul51 reverse transcriptase (RT) protein.  
 XX  
 KW Regulatory gene; accessory gene; HIV; human immunodeficiency virus;  
 KW vaccine; infection; gene therapy; reverse transcriptase; RT; enzyme.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 PN WO2003037919-A2.  
 XX  
 PD 08-MAY-2003.  
 XX  
 PF 31-OCT-2002; 2002WO-1B004550.  
 XX  
 PR 31-OCT-2001; 2001ZA-00008978.  
 XX  
 PA (SAME-) SOUTH AFRICAN MEDICAL RES. COUNCIL.  
 PA (UYCA-) UNIV CAPE TOWN.  
 XX  
 PI Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;  
 XX





1963 AAGAGCCCATCATCGCGCCGAGACCTTCTACGTGGACGGCGCCGCAACCCGAGACC 2022  
 581 LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600  
 2023 AAGATCGGCAAGCCGCTACGTACCGACCGCGCGCGGAGATCGTGACCTGACC 2082  
 601 LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620  
 2083 GAGACCAACCAACGAGACCGAGCTGACCGACCGCGCGCGGAGATCGTGACCGGC 2142  
 621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640  
 2143 AGCAGGTGAACATCGTGACCGACGACGACGACGACGACGACGACGACGACGAC 2202  
 641 SerGluValAsnIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660  
 2203 GACAAGAGCGAGCGAGCTGTGTGAACACGATCATCGAGCAGCTGTGTGAAGAGAG 2262  
 661 AspArgSerGluSerGluLeuValAsnGlnIleIleGlnLeuIleLysLysGluArg 680  
 2263 GTGTACTGTGCTGGTGGCGCGCCACGAGGATCGCGCGCGACGAGCAGATCGACAG 2322  
 681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700  
 2323 CTGTGTGAGCAAGGCGCATCCGCAAGGTCTCTCTGTGACGCGCATCGAT 2370  
 701 LeuValSerArgGlyLeuArgLysValLeuValLeuAspGlyIleAsp 716

RESULT 8  
 AAB86169  
 ID AAB86169 standard; protein; 3025 AA.  
 AC AAB86169;  
 XX  
 XX  
 11-SEP-2003 (revised)  
 21-AUG-2001 (first entry)  
 XX  
 XX  
 HIV-1 subtype C protein fragment #1.  
 XX  
 XX  
 Infection; diagnosis; human; humoral immune response; antiviral;  
 cellular immune response; vaccine; treatment; gene therapy.  
 XX  
 XX  
 Human immunodeficiency virus 1.  
 XX  
 XX  
 Key Location/Qualifiers  
 FT Misc-difference 1..3025  
 FT /label= Xaa  
 FT /note= "Xaa represents a stop codon"  
 XX  
 DE10056747-Al.  
 XX  
 31-MAY-2001.  
 XX  
 16-NOV-2000; 2000DE-01056747.  
 XX  
 16-NOV-1999; 99DE-01055089.  
 XX  
 (SHAO/) SHAO Y.  
 PA (GENE-) GENEART GMBH GES ANGEWANDTE BIOTECHNOLOG.  
 XX  
 Wagner R, Wolf H, Shao Y, Graf M;  
 XX  
 WPI; 2001-336417/36.  
 DR N-PSDB; AAH20870.  
 XX  
 New nucleic acid sequences from a human immune deficiency virus  
 PT intersubtype, useful for treatment, prevention and diagnosis of  
 PT infection.  
 XX  
 Disclosure; Fig 8A-O; 48pp; German.  
 PS  
 XX  
 This invention describes a novel polynucleotide isolated from human  
 CC immunodeficiency virus type 1 subtype C/B' which can be used for the

induction of specific humoral and cellular immune responses. (I) and  
 polypeptides (II) encoded by them, are useful in pharmaceuticals,  
 vaccines and diagnostic agents, particularly for treatment or prevention  
 of human immune deficiency virus-1 (HIV-1) infections, also for rational  
 design of test or therapeutic reagents, or gene therapy vectors.  
 CC Polypeptides, especially antibodies, specifically directed against (II)  
 CC are similarly useful as pharmaceutical and diagnostic agents. (I) are  
 CC specific for intersubtype C/B' of HIV-1 so are useful in regions  
 CC (particularly China and South-East Asia) where this subtype is prevalent.  
 CC the products of the invention have antiviral activity. This sequence  
 CC represents a protein encoded by the HIV-1 subtype C genome described in  
 CC the method of the invention. (Updated on 11-SEP-2003 to standardise OS  
 CC field)  
 XX  
 SQ Sequence 3025 AA;  
 Alignment Scores: 8.62e-203 Length: 3025  
 Pred. No.: 3601.00 Matches: 686  
 Score: 92.15% Conservative: 42  
 Percent Similarity: 86.84% Mismatches: 55  
 Best Local Similarity: 78.57% Indels: 7  
 Query Match: 4  
 DB: 4  
 US-09-610-313B-30 (1-2469) x AAB86169 (1-3025)  
 QY 15 GGCAGGAGCCATGAGCCAGCCAGCCACATCTCTGATCGAGCGAGCAACTTCAA 74  
 Db 416 GlyArgGlyAsnGluProAsnLysGln---CysHisThrAspAlaGluLysGlnPhe\*\*\* 434  
 QY 75 GGGCCCAAGAGCGATCATCAAGTCTTCAACTCGCGGAGAGGGCCACATCGCCCGCAA 134  
 Db 435 ArgLeu\*\*\*LysAsnCys\*\*\*MetPheGlnLeuTrpGlnGlyArgAlaHisSerGlnLys 454  
 QY 135 CTGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGCCACAGATGAA 194  
 Db 455 LeuGlnGlyPro\*\*\*GluLysGlyLeuLeuGluMetTrpLysArgArgThrProAsnGlu 474  
 QY 195 GGNCTGACGAGCGCGCCGACCTCTTCCGCGAGGAGCTGCGCTTCCCGAGGCA 254  
 Db 475 ArgLeuTyr\*\*\*GluThrGlyGln-PhePheArgGluAsnLeuAlaLeuProGlnGlyAr 494  
 QY 255 GGCAGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCGCCAGCCCGAGCTGCA 314  
 Db 494 GAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerProThrArgGlyGluLeuG 514  
 QY 315 GGTG-----CGCGCGCAACACCCCGCGAGGCGCGCGCGCGAGCGCCGAGCCCT 368  
 Db 514 nValTrpGlyArgAspAsnAsnSerIleSerGluAlaGlyAlaAsnArgGlnGlyThrI 534  
 QY 369 G-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCTCGTGAGCATCAAGGTGG 422  
 Db 534 eSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleG 554  
 QY 423 CGGCAGATCAAGAGCGCCCTGTGTGACACACCGCGCGCGAGCACACCGTGTGTGAGGAGAT 482  
 Db 554 yGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaGlyAspThrValLeuGluaspLe 574  
 QY 483 GAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAGGT 542  
 Db 574 uAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyPheIleLysVa 594  
 QY 543 GCGCAGTAGCAGACCATCTGATCAGATCTGGGCGAGGAGGAGGAGGAGGAGGAGGAG 602  
 Db 594 lArgGlnTyrGluGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThrValLe 614  
 QY 603 GATCGCGCCCGCCCGTGAACATCATCGCGCGCCCAACATCTGACCCAGCTGGGCTGCAC 662  
 Db 614 uValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysTh 634  
 QY 663 CTGGAATCTCCCATCAGCCCCATCAGACCGCGTGGCGCGTGAAGTGAAGCGCGCATGGA 722  
 Db 634 rLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAs 654

QY 723 CGGCCCCAAGGTGAAGCAGTGGCCCTTACCGAGGAGAGATCAAGGCCCTGACGGCCAT 782  
Db 654 pGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaI 674  
QY 783 CTGGAGGAGATGGAGAGAGGAGGAGATCACCAAGATCGGCCCGGAGAACCCCTACAA 842  
Db 674 eCyAspGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAs 694  
QY 843 CACCCCGTGTTCGCCATCAAGAAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGACTT 902  
Db 694 nThrProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPh 714  
QY 903 CCGGAGCTGAACAGGACCCAGGACTTCTGGAGGTGAGTGGCATCCGCCACCC 962  
Db 714 eArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734  
QY 963 CGCCGGCTGAAGAAGAGAGCGTGACCGTGTGGAGCTGGCGAGCGCTACTTTCAG 1022  
Db 734 oAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754  
QY 1023 CGTCCCTCGGAGGAGGATTCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAA 1082  
Db 754 rIleProLeuTyrGluAspPheArgLysTyrThrAlaPheThrIleProSerArgAsnAs 774  
QY 1083 CGAGACCCCGGATCCGCTACCAAGTACACGTCGTGCTGGCGTGGGAGGAGCGGCC 1142  
Db 774 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerLe 794  
QY 1143 CAGCATCTTCAGAGCAGCATGACCAAGATCCTCGAGCCCTTCCGCGCCCGCACCCCGA 1202  
Db 794 uAlaIlePheGlnSerSerMetThrLysThrLeuGluProPheArgLysGlnAsnProGl 814  
QY 1203 GATCGTATCTACAGTACATGAGCAGCTGTAGTGGGAGCGACCTGGAGATCGGCCA 1262  
Db 814 yIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGl 834  
QY 1263 GCACCGGCCAAGATCGAGGAGTGGCGAAGCAGTCTGCTGGCTGGGCTTCACACCCC 1322  
Db 834 nHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrPr 854  
QY 1323 CGACAAAGCAGCAGAGGAGGCCCTTCTGTGATGGCTACGAGTGCACCCCGA 1382  
Db 854 oAspLysLysHis---LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAs 873  
QY 1383 CAAGTGACCGTGCAGGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACAT 1442  
Db 873 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerTrpThrValAsnAspIl 893  
QY 1443 CCAGAAGCTGGTGGCAAGCTGAATCGGCGAGCCAGATCTACCCCGCATCAAGGTGCG 1502  
Db 893 eGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValAr 913  
QY 1503 CCAGCTGTGACAGTGTGCGCGGCGCCAGGCCCTGACCGACATCTGCGCCCTGACCGA 1562  
Db 913 gGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGl 933  
QY 1563 GGAGCGCGAGTGGCTGGCGGAGAACCGGAGATCTGCGCGAGCCGCTGACGCGCT 1622  
Db 933 uGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLysGluProValHisGlyVa 953  
QY 1623 GTACTACGACCCAGCAGGACCTGGTGGCGAGATCCAGAAGCAGGCGCCACGACGATG 1682  
Db 953 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973  
QY 1683 GACCTACAGATCTACAGAGGCCCTTCAAGAACCTGAAGCCGAGAGTACGCGCAAGAT 1742  
Db 973 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993  
QY 1743 GCGCAGCGCCACACCAAGCAGTGAAGCAGCTCACCGAGCGCTGAGAGATCGCCAT 1802  
Db 993 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

QY 1803 GGAGAGCATCGTGTATCTGGGGCAAGACCCCAAGTTCGCTGCGCATCCAGAACGAGAC 1862  
Db 1013 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033  
QY 1863 CTGGGAGACCTGGTGGAGCCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGT 1922  
Db 1033 rTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053  
QY 1923 GAACACCCCGCTGTGTGAAGCTGTGTACAGCTGGAGAAAGAGCCCATCATCGGCGC 1982  
Db 1053 lAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysAspProIleValGlyVa 1073  
QY 1983 CGAGACCTTCTACCTGAGCGCGCGCCCAACCGCGAGACCAAGATCCGCGAGGCGCGCTA 2042  
Db 1073 lGluThrPheTyrValAspGlyAlaAlaAsnArgLysThrLysIleGlyLysAlaGlyTy 1093  
QY 2043 CGTACCGACCGCGCGCGCAGAGATCGTGTGAGCTGACCGAGACCAACCAACCAAGAC 2102  
Db 1093 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113  
QY 2103 CGAGCTCGAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGTGAACATCGTAC 2162  
Db 1113 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133  
QY 2163 CGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGCCCGCCAGAGCGAGCGAGCT 2222  
Db 1133 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153  
QY 2223 GGTCAACCCAGATCATCCAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCC 2282  
Db 1153 uValAsnGlnIleIleGluLeuMetLysLysGluArgValTyrLeuSerTrpValPr 1173  
QY 2283 CGCCCAACAGGCGCATCGCGCGCAACCGAGCAGATCGACAAGCTGGTGAGCAAGGCGCATCG 2342  
Db 1173 oAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleAr 1193  
QY 2343 CAAGTGTCTTCTGCGACCGCATCGAT 2370  
Db 1193 gLysValLeuPheLeuAspGlyIleAsp 1202  
RESULT 9  
AAB69282  
ID AAB69282 standard; protein; 1000 AA.  
XX  
AC AAB69282;  
XX  
XX 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX  
XX HIV-1 non-subtype B clone 92RW009-6 pol protein.  
DE  
XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
KW vif; vpr; tat; rev; nef; vaccine.  
XX  
OS Human immunodeficiency virus 1.  
XX  
FN WO200026416-A1.  
XX  
PD 11-MAY-2000.  
XX  
PF 25-OCT-1999; 99WO-US024837.  
XX  
PR 02-NOV-1998; 98US-00184418.  
XX  
PA (UABR-) UAB RES FOUND.  
XX  
XX Hahn BH, Shaw GM, Gao F;  
PI  
XX  
DR WPI; 2000-365651/31.  
XX  
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.

XX Claim 41; Fig 15; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 1000 AA;

SQ

Alignment Scores:

Pred. No.:	3 07e-200	Length:	1000
Score:	3555.50	Matches:	664
Percent Similarity:	96.52%	Conservative:	29
Best Local Similarity:	92.48%	Mismatches:	24
Query Match:	77.58%	Indels:	1
DB:	3	Gaps:	1

US-09-610-313B-30 (1-2469) x AAB69282 (1-1000)

QY 220 TTCTTCCGCGAGGACCTGGCTTCCCTCCCGAGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279

DB 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20

QY 280 AACCGGCGCAACAGCCACCACCGCGGAGCTGAGTGGCGCGC---GACAACCCCGCGC 336

DB 21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyGlyArgAspSerLeuSer 40

QY 337 AGCGAGCGCGCGCGAGCCAGGCGCACCTCAACTTCCCTCCCGAGATCACCTGTGGCGAG 396

DB 41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnLeuLeuTrpGln 60

QY 397 CGCCCTCGTGTGACATCAAGTGGCGCGCGAGATCAAGAGAGCCCTGTGGACACCGGC 456

DB 61 ArgProLeuValThrValIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80

QY 457 GCGAGACACCTGTGTGGAGGAGATGAGCTCCCGGCAAGTGGAGCCCAAGATGATC 516

DB 81 AlaAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIle 100

QY 517 GCGGAGATCGCGGCTTCAATCAAGTGGCGCGAGTACGACAGATCCTGTGATCGAGATCTGC 576

DB 101 GlyGlyIleGlyPheIleLysValLysGlnIleThrAspGlnLeuLeuIleGluIleCys 120

QY 577 GCGAAGAGCCATCGGCACCGCTGTGATCGGCGCCACCCCGTGAAACATCATCGGCGCG 636

DB 121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleIleGlyArg 140

QY 637 AACATGTGACCGAGTGGGCTGACCTGAATCTTCCCATCAGCCCGCATCGAGACCGTG 696

DB 141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160

QY 697 CCGGTGAAGTGAAGCCCGCATCGGCGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAG 756

DB 161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnIleProLeuThrGlu 180

QY 757 GAGAAGATCAAGCCCTTGACCGGCATCTGCGAGGAGATGAGAGAGGAGGCGCAAGATCAC 816

DB 181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluIleLysIleSer 200

QY 817 AGATCGGCGCGAGAACCCCTTCAACACCCCGTGTTCGCCATCAAGAAAGAGGACAGC 876

DB 201 LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 220

QY 877 ACCAAGTGGCGAGCTGGTGGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGG 936

DB 221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240

QY 937 GAGGTGACGCTGGGCGATCCCGCCCGCGCTGGAAGAAGAAGAGCGTGGACCGTG 996

DB 241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrVal 260

QY 997 CTGGAGCTGGCGGACCGCCTTACGCTGCGCCCTCGAGCGAGACTTCCGCAAGTACACC 1056

DB 261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280

QY 1057 GCCTTCCACCATCCCGAGCATCAACAAAGAGACCCCGGATCCGCTACCAAGTCAACAGTG 1116

DB 281 AlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 300

QY 1117 CTGCCCCAGGGCTGGAAGGCGACCCCGACGATCTTCCAGAGCAGCATGACCAAGATCCTG 1176

DB 301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320

QY 1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGTATCTACAGTACATGAGACCGCTGTAC 1236

DB 321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspLeuTyr 340

QY 1237 GTGGCGAGCGACCTGGAGATCGGCCAGCACCGGCCCAAGATCGAGAGCTGCGCAAGCAC 1296

DB 341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuArgGluHis 360

QY 1297 CTGCTCGGCTGGGGCTTCAACCCCGCGAGAGCACCAAGAGAGCGCCCTTCTCTG 1356

DB 361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 380

QY 1357 TGGATCGGCTTACGAGCTGCACCCCGCAAGTGCAGCTGCAGCCCATCGAGCTGCCCGAG 1416

DB 381 TrpMetGlyTyrGlnLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400

QY 1417 AAGGAGAGCTGGACCGTGAACGACATCCAGAACCTGCTGGCGAAGCTGAATCTGGCGCAG 1476

DB 401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420

QY 1477 CAGATCTACCCCGGATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGCGCGCCAGGCC 1536

DB 421 GlnIleTyrProGlyValLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAla 440

QY 1537 CTGACCGACATCTGCTGCGCCCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAG 1596

DB 441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 460

QY 1597 ATCTCTCGCGAGCCCGTGCACCGCGCTGTACTACGACCCCGAGAGAGCTGTGGTGGCGAG 1656

DB 461 IleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGlu 480

QY 1657 ATCCAGAGCAGGCGCCAGCAGCTGACCTACAGATCTACAGAGCGCCCTTCAAGAAC 1716

DB 481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500

QY 1717 CTCAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCGCACCAACAGCAGCTGAAGCAGCTG 1776

DB 501 LeuLysThrGlyLysTyrAlaLysArgThrAlaHisThrAsnAspValLysGlnLeu 520

QY 1777 ACCGAGCGCTGCAGAAGATCGCCATCGAGAGATCGTGTATCTGGGCGAAGACCCCAAG 1836

DB 521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleIleTrpGlyLysThrProLys 540

QY 1837 TTCGCGCTGCCATCCAGAAAGAGACCTGGGAGACCTGTGGAGCCGACCTACTGCGAGGCC 1896

DB 541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAla 560

QY 1897 ACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTGTACCG 1956

DB 561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580

QY 1957 CTGGAAGAGGAGCCCATCATCTCGCGCGCGAGACCTTCTACGTGAGCGCGCGCAACCGC 2016

DB 581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

QY 2017 GAGACCAAGATCGGCAAGCGCGCTACGTGACCGCGCGCGCGCGAGAGATCGTGACG 2076

Db 601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620  
 Qy 2077 CTGACCCGAGACCAACCAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTCGAGGAC 2136  
 Db 621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640  
 Qy 2137 AGCGGACGAGGTGAACATGTCACCGACAGCCAGTACGGCCCTGGGCATCATCCAGGCC 2196  
 Db 641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleGlnAla 660  
 Qy 2197 CAGCCCGACAAAGACGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTCATCAAGAAG 2256  
 Db 661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleGlnLeuIleLysLys 680  
 Qy 2257 GAGAAGGTGTACTGAGCTGGGTGCCGCCACCAAGGCCATCGCGGCAACGACGATC 2316  
 Db 681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 700  
 Qy 2317 GACAAGCTGTGAGCAAGGGCATCGCAAGGTGCTGTCTCGAGCGGCATCGAT 2370  
 Db 701 AsplysLeuValSerSerGlyIleArgArgValLeuPheLeuAspGlyIleAsp 718  
 RESULT 10  
 AAW72993  
 ID AAW72993 standard; protein; 1002 AA.  
 XX  
 AC AAW72993;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 15-FEB-1999 (first entry)  
 XX  
 DE HIV isolate LAV.MAL pol protein.  
 XX  
 KW LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;  
 KW pol.  
 XX  
 OS Human T-lymphotropic virus.  
 XX  
 PN US5824482-A.  
 XX  
 PD 20-OCT-1998.  
 XX  
 PF 06-JUN-1995; 95US-00471474.  
 XX  
 PR 23-JUN-1986; 86FR-00040138.  
 PR 13-APR-1987; 87US-00038330.  
 PR 19-FEB-1991; 91US-00656797.  
 PR 10-DEC-1992; 92US-00988530.  
 PR 18-NOV-1993; 93US-00154397.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Wain-Hobson S, Sonigo P, Alizon M, Montagnier L;  
 XX  
 DR WPI; 1998-582548/49.  
 DR N-PSDB; AAV63467.  
 XX  
 XX Human immunodeficiency virus isolate LAV(MAL) - and method for detecting  
 PT anti-HIV antibodies.  
 PT  
 PS Disclosure; Fig 7B-E; 47pp; English.  
 XX  
 CC This is the amino acid sequence of the pol protein of lymphadenopathy-  
 CC associated virus LAV.MAL (CNCM I-641), a new virus isolate from Zaïre  
 CC that is responsible for diseases clinically related to AIDS. The sequence  
 CC was deduced from an open reading frame (ORF) of the LAV.MAL genome (see  
 CC AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of  
 CC the envelope glycoprotein can be used as antigens in a claimed method for  
 CC the in vitro detection of an antibody directed against LAV. The method is  
 CC useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in  
 CC patients, asymptomatic carriers and in blood-related products. LAV viral  
 CC antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct  
 CC PR field.)

XX SQ Sequence 1002 AA;  
 Alignment Scores:  
 Pred. No.: 3,49e-199 Length: 1002  
 Score: 3537.50 Matches: 657  
 Percent Similarity: 96.67% Conservative: 39  
 Best Local Similarity: 91.25% Mismatches: 21  
 Query Match: 77.19% Indels: 3  
 Gaps: 2  
 DB:  
 US-09-610-313B-30 (1-2469) x AAW72993 (1-1002)  
 Qy 220 TTCTTCCGCGAGACCTGGCTTCCCTCCAGGGCAAGGCCCGAGATTCCCGAGCGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGlnGln 20  
 Qy 280 AACCGCGCCACACAGCCACCGCGCGAGCTGCAGGTG---CGCGGCGACACCCCGCGC 336  
 Db 21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40  
 Qy 337 AGCGAGCGCGCGCGCGAGCGCGAGGCG---ACCTTGAACCTTCCCGCGAGATCACCCCTG 390  
 Db 41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPhePheProGlnIleThrLeu 60  
 Qy 391 TGGCAGCGCCCTGGTGAGCATCAAGTGGCGCGCGAGATCAAGAGGCGCCCTGCTGGAC 450  
 Db 61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuAsp 80  
 Qy 451 ACCGCGCGCGCGAGCAGCGCTGCTGAGGAGATGAGCTGCGCGGCGAGTGGAGGCCCAAG 510  
 Db 81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100  
 Qy 511 ATGATCGCGCGCATCGCGCTTTCATCAAGTGGCGCGAGTACGACCGAGATCCTGATCCAG 570  
 Db 101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120  
 Qy 571 ATCTGCGCGCAAGAGCCATCGGCATCGCTGCTGATCGCGCCCGCCCGCTGAAACATCATC 630  
 Db 121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140  
 Qy 631 GGCGCGACATGCTGACCGCGCTGGCTGCGACCTGCACTTCCCGCATCAGCCCGCATCAG 690  
 Db 141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
 Qy 691 ACCGTGCGCGTGAAGCTGAAGCCGCGATGACCGCGCCCAAGGTGAAGCAGTGGCGCCCTG 750  
 Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180  
 Qy 751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCCATCTTCGAGGAGATGGAGAGGAGGCGCAAG 810  
 Db 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200  
 Qy 811 ATCAACGAAGATCGCGCGCGAGAACCCCTACACACCCCGCTGTCGCGCATCAAGAGAG 870  
 Db 201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220  
 Qy 871 GACAGCACCAAGTGGCGCAAGCTGGTGCAGCTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930  
 Db 221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240  
 Qy 931 TTCTGGGAGGTGCGAGCTGGGCATCCCCACCCCGCGCTGAAAGAGAAAGAGAGCGCTG 990  
 Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
 Qy 991 ACCGTGCTGAGCTGGCGGAGCCCTACTTACGCTGCGCCCTGGAGCAGGAGCTTCGCGCAG 1050  
 Db 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280  
 Qy 1051 TACACGCGCTTACCATCCCGAGCATCAACAAGCAGACCCCGCGCATCCGCTACAGTAC 1110  
 Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300



## Alignment Scores:

Pred. No.: 3.59e-199 Length: 854  
Score: 3537.00 Matches: 667  
Percent Similarity: 94.44% Conservative: 30  
Best Local Similarity: 90.38% Mismatches: 30  
Query Match: 77.18% Indels: 11  
DB: 5 Gaps: 2

US-09-610-313B-30 (1-2469) x AAM48949 (1-854)

```
QY 171 CGGCAAGGAGGGCCACCATGATGAAGGCTGCACCGAGCGCCAGGCGCAACTTCTT----- 224
DB 118 ArgGlnGlyPheProSerHisAspValValLysArgArgProValProSerLeuHisAla 137
QY 225 -----CCGCGAGGACTTGGCC-----TTCCGCCAGGGCAAGGCCCG 260
DB 138 CysArgSerThrLeuGluAspProArgValPro-SerSerPheProGlnGlyProAlaAr 157
QY 261 CGAGTTCCTCCAGCGAGCAGAAACCGCGCCAAACAGCCCAACAGCCCGAGCTGCAGGTGCG 320
DB 157 gGlnPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeuGlnValAr 177
QY 321 CGGCGAACCCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
DB 177 gArgAspAsnProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGl 197
QY 381 GATCACCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGGTGGCGCGCCAGATCAAGGAGGC 440
DB 197 nileThrLeuTrpGlnArgProLeuValSerileLysileGlyGlnThrArgGluAl 217
QY 441 CCTGCTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
DB 217 aLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTr 237
QY 501 GAACCCCAAGATGATCGCGCGCATCGCGCGCTTATCATAGGTGGCGCGCATGACGACCAT 560
DB 237 pLysProLysMetileGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIl 257
QY 561 CCTGATCGAGATCTCGGCAAGAGGCGCATCGGACCGTGTGATCGGCGCCACCCCGCT 620
DB 257 eLeuileGluileCysGlyLysLysAlaileGlyThrValLeuValGlyProThrProva 277
QY 621 GAACATCATCGCGCGCAACATGTGACCCAGCTGGCTGACCCCTGACCTTCCCGCATCAG 680
DB 277 lAsnileileGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProLysE 297
QY 681 CCCCATCGAGACCTGCGCGTGAAGCTGAAGCCCGGCGATGACCGCCCAAGGTGAAGCA 740
DB 297 rProileGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGl 317
QY 741 GTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAGATGGAGAA 800
DB 317 nTrpProLeuThrGluValLysIleLysAlaLeuThrAlaileCysGluGluMetGluLy 337
QY 801 GGAGGCAAGATCACCAAGATCGGCGCGCGGAAACCCCTACAAACCCCGCTGTTCCGCAT 860
DB 337 sGluGlyLysileThrLysileGlyProGluAsnProLysrAsnThrProLysPheAlaIl 357
QY 861 CAAGAAGAGACAGCACCAGTGGCGGCAAGCTGGTGACTTCCGCGAGCTGAACAAGCG 920
DB 357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysAr 377
QY 921 CACCCAGGACTTCGCGAGGTGAGCTGGGCATCCCGCCCGCGCGCTGGAAGAGAA 980
DB 377 sThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLy 397
QY 981 GAAGAGCGTCAACCGTGTGAGCGGCGCGCGCTACTTCAAGCTGGCGCTGGACGAGGA 1040
DB 397 sLysSerValThrValLeuAspValGlyAspAlaTyrrPheSerValProLeuAspGluGl 417
QY 1041 CTTCCGCAAGTACACCGCTTTCACATCCCGCATCAACAGAGACCCCGCGCATCCG 1100
```

```
DB 417 yPheArgLysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleAr 437
QY 1101 CTACAGTACAAACGTGTGCTGCCCGCAGGGCTGAAGGCGAGCCAGCATCTTCCAGAGCAG 1160
DB 437 gTyrrGlnTyrrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSe 457
QY 1161 CATGACCAAGATCCTGAGAGCCCTTCCGGCGCGCGCAACCCCGAGATCGTGATCTACAGTA 1220
DB 457 rMetThrLysileLeuGluProPheArgAlaLysAsnProGluIleValIleTyrrGlnTy 477
QY 1221 CATGCGACGACCTGTACGTGGGCGAGCCTCGAGATCGCGCAGCAGCAGCCGCGCAAGATCGA 1280
DB 477 rMetAlaAlaLeuTyrrValGlySerAspLeuGluileGlyGlnHisArgAlaLysileGl 497
QY 1281 GGAGCTCGCGCAAGCACCTGTGCTGGGGCTTCCACACCCCGCGCAAGAGCAGCAGAA 1340
DB 497 uGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLy 517
QY 1341 GGAGCCCGCTTCTCTGTGATGGCTAGGCTAGCAGCTGCACCCCGCAACAGTGGCGCGAGCC 1400
DB 517 sGluProProPheLeuTrpMetGlyTyrrGluLeuHisProAspLysTrpThrValGlnPr 537
QY 1401 CATCGAGCTGCCGAGAGGAGAGAGCTGCACCGTCGAAACGACATCCAGAAAGCTGGTGGCAA 1460
DB 537 oileGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLy 557
QY 1461 GCTGAACCTGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGCAGCTGTGCAAGCTGCT 1520
DB 557 sLeuAsnTrpThrSerGlnileTyrrProGlyLysValArgGlnLeuCysLysLeuLe 577
QY 1521 GCGCGCGCGCAAGCCCTGACCGATCGTGCCTGACCCGAGGAGCGCGAGCTGGAGCT 1580
DB 577 uArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLe 597
QY 1581 GGCGCGAAGCGCGAGATCCTCGCGAGCGCGTGCACCGCGTGTACTACGACCCCGCAGCAA 1640
DB 597 uAlaGluAsnArgGluLeuLysGluProValHisGlyValTyrrAspProSerLy 617
QY 1641 GGACCTGCTGCCGAGATCCAGAAAGAGGCGCGCAGCAGCTGAGTGGACCTACAGATCTACCA 1700
DB 617 sAspLeuileAlaGluileGlnLysGlnGlyAspAspGlnTrpThrTyrrGlnileTyrrGl 637
QY 1701 GGAGCCCTTCAAGAACTGAGACCGGCAAGTACGCCAAGATGGCGACCGCCCGCACACCAA 1760
DB 637 nGluProPheLysAsnLeuLysThrGlyLysTyrrAlaLysArgArgThrHisThrAs 657
QY 1761 CGAGCTGAGCAGCTGACCGCGCGCTGCAGAGATCGCCATGAGAGCATCGTGATCTG 1820
DB 657 nAspValLysGlnLeuThrGluAlaValGlnLysileSerLeuGluSerileValThrTr 677
QY 1821 GGCAAGACCCCGCAAGTTCGCGCTGCCCATCCAGAAAGAGAGACCTGGGAGACCTGTGGAC 1880
DB 677 pGlyLysThrProLysPheArgLeuProileGlnLysGluThrTrpGluileTrpTrpTh 697
QY 1881 CGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCGCTGGT 1940
DB 697 rAspTyrrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVa 717
QY 1941 GAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGGA 2000
DB 717 lLysLeuTrpTyrrGlnLeuGluLysGluProileAlaGlyAlaGluThrPheTyrrValAs 737
QY 2001 CCGCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGCGGCGCG 2060
DB 737 pGlyAlaAlaAsnArgGluThrLysileGlyLysAlaGlyTyrrValThrAspArgGlyAr 757
QY 2061 GCAGAAGATCGTGAGCTGACCGGAGACCAACCAACAGAAAGCCAGCTGCGAGGCCATCCA 2120
DB 757 gGlnLysileValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaileGl 777
QY 2121 GCTGGCGCTGCAGACCGCGCAGCGAGGTGCAACATCGTGACCGCAGCGCAGTACGCCCT 2180
DB 777 nLeuAlaLeuGlnAspSerGluSerGluValAsnileValThrAspSerGlnTyrrAlaLe 797
```





```
Qy 1408 CTGCCCGAGAGGAGCTGACCGTGAACCATCCAGAGCTGGTGGCGCAAGCTGAAC 1467
Dy 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
Qy 1468 TGGGCCAGCAGATCTACCCCGCATCAAGGTGGCCAGCTGTCAAGCTCTCGCGGCG 1527
Dy 421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
Qy 1528 GCCAAGGCCCTGACCGCATCGTCCCTGACCCAGGAGGCGGAGCTGGAGCTGCCCGAG 1587
Dy 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
Qy 1588 AACCGCAGATCTCGCGAGCCCGTGCACCGCTGTACTAGCAGCCCGCAGCAAGACTG 1647
Dy 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
Qy 1648 GTGCCCGCAGATCCAGAGCAGGCGCACACAGTGGACCTACCATGATCTACCGAGCGCC 1707
Dy 481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
Qy 1708 TTCAAGAACCTGAAAGACCGGCAATAGCTAGCCAAAGATGGCCAGCCCAACCAAGCGTG 1767
Dy 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
Qy 1768 AAGCAGCTGACCGAGCGCTGCAGAGATCCCATGGAGAGCATCGTGATCTGGGCAAG 1827
Dy 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540
Qy 1828 ACCCCCAAGTCCCGCTGCCCATCCAGAGAGACCTGGGAGACCTGGTGACCGACTAC 1887
Dy 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
Qy 1888 TGGCAGCGCACTCGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTCAAGCTG 1947
Dy 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
Qy 1948 TGTACCACTGGAGAGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGGCGCC 2007
Dy 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
Qy 2008 GCCAACCGCAGACCATGATCGGCAAGCCGGCTACGTGACCGACCGCGGCGCGCAGAG 2067
Dy 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
Qy 2068 ATCGTGAGCTGACCGCAGACCAACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGCC 2127
Dy 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
Qy 2128 CTGCAGGACAGCGCAGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Dy 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Qy 2188 ATCCAGGCCAGCCGACAGAGCGAGCGAGCTGGTGAACCATGATCATCGACAGCTG 2247
Dy 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
Qy 2248 ATCAAGAGAGAGAGGTGTACTCGAGCTGGGTGGCCGCCCAAGAGGCATCGCGGCGAAC 2307
Dy 681 IleLysLysGlnLysValTyrLeuAlaIleTrpValProAlaHisLysGlyIleGlyLysn 700
Qy 2308 GAGCAGATCGACAGCTGGTGAGCAAGGCGATCCGCAAGGTGTCTTCTGGAGCGGCATC 2367
Dy 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
Qy 2368 GAT 2370
Dy 721 Asp 721
RESULT 13
ADN36413
ID ADN36413 standard; protein; 1003 AA.
XX
```

```
AC ADN36413;
XX 15-JUL-2004 (first entry)
DT XX HIV protein for anti-HIV vaccine.
DE XX anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX XX Human immunodeficiency virus 1.
OS XX WO2004035006-A2.
PN XX 29-APR-2004.
PD XX 17-OCT-2003; 2003WO-US033112.
XX XX 18-OCT-2002; 2002US-0419465P.
PR XX (AARO-) AARON DIAMOND AIDS RES CENT.
XX XX Huang Y, Ho DD, Chen Z;
XX XX WPI: 2004-348328/32.
XX XX N-PSDB; ADN36412.
XX XX Nucleic acid vector comprising at least one HIV sequence operably linked
PT to a promoter and encoding a protein that does not assemble into viral
PT particles, useful in immunizing a subject against HIV infection.
XX XX Disclosure; SEQ ID NO 27; 166pp; English.
XX XX The invention relates to a nucleic acid vector comprising at least one
CC HIV sequence operably linked to a promoter and encoding a protein that
CC does not assemble into viral particles. The nucleic acid vector is useful
CC in immunizing a subject against HIV infection. This sequence corresponds
CC to a nucleic acid used in the invention.
XX XX
```

SQ Sequence 1003 AA;

Alignment Scores:					
Pred. No.:	4.9e-199	Length:	1003		
Score:	3535.00	Matches:	659		
Percent Similarity:	96.12%	Conservative:	34		
Best Local Similarity:	91.40%	Mismatches:	24		
Query Match:	77.13%	Indels:	4		
DB:	8	Gaps:	2		

US-09-610-313B-30 (1-2469) x ADN36413 (1-1003)

```
Qy 220 TTCTTCCGAGGACCTGGCTTCCCGGCAAGGCCCGCGAGTTCCCGAGCGAGCAG 279
Dy 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
Qy 280 AACCGGCCAACACCCCGCCAGCGCGAGCTGCAGGTG-----CGCGGCGACACCCC 333
Dy 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSer 40
Qy 334 CGCAGCGCAGCGCGCGCGAGCGCGAGCGCGACCGCTG-----AACTTCCCGCAGATCAC 387
Dy 41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
Qy 388 CTGTGGCAGCGCCCGCTGGTGAGCATCAAGTGGCGCGCCAGATCAAGAGCGCCCTGTG 447
Dy 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
Qy 448 GACACCGCGCGCGAGCAGCACCGCTGCGAGAGATGAGCTGCCCGGCAAGTGAAGCCC 507
Dy 81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100
Qy 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCGCAGTACGACCATCATC 567
Dy 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
```

```
QY 568 GAGATCTGCGCGCAAGGCCATCGGCACCGTGTGATCGGCCCCACCCCGTGAACATC 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGGCGCAACATGTCAGCCAGCTGGGTGTCACCTGAATTCCTCCCATCAGCCCATC 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGAGCTGAGCCCGGATGAGCCGATCGAGCGGCCCAAGGTGAGCAGTGC 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGCGCCCTGACCGCCATCTGCGAGGAGATGAGAAAGAGGCG 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCCCGGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAG 867
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGCAGACGACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGACAGCGCACCCAG 927
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGGAGGTGACGTGGGCATCCCCACCCCGCGCTGGAAGAAAGAAAGAGC 987
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTACCGCTGTCAGCTGGCGCGCGCTACTTTCAGCGTGCCTTGACGAGGACTTCGCG 1047
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280
QY 1048 AAGTACACCGCTTCACCATCCCGCAGCATCAACAGCAGACCCCGCATCCGCTACCAG 1107
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGTGCCCGCAGGCTCGAAGGCGAGCCCGCAGCATCTTCAGAGCAGCATGACC 1167
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCCTGGAGCCCTTCGCGCGCCCGCAACCCCGAGATCGTGATCTACAGTACATGGAC 1227
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY 1228 GACTGTAGTGGCGAGCGCTGAGATCGGCAGCAGCCCGCCAGATCGAGGAGCTG 1287
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1288 CGCAAGCACCTGTGCTGGGCTTCACCAACCCCGCAGCAAGAACAGCAGAGAGAGCC 1347
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380
QY 1348 CCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAG 1407
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1408 CTGCCCGAGAGAGAGCTGGACCGTGCAAGCATCCAGATCCAGAGCTGGTGGCAGCTGAAC 1467
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1468 TGGGCGCAGCAGATCTACCCCGCGCATCAAGGTGCGCGAGCTGTGCAAGCTGCTCGCGGC 1527
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1528 GCCAAGCCCTGACCGACATCGTGCCCTGACCGAGGAGCCGAGCTGGAGCTGGCGGAG 1587
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1588 AACCGCGAGATCTGCGCGAGCCGCTGACGGGTGTACTACGACCCCGCAGCAGGACCTG 1647
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1648 GTGGCCGAGATCCAGAAGCAGGGCCACGACAGTGGACCTTACCAGATCTACCAGGAGCC 1707
```

```
Db ::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
QY 1708 TTCAAAGAACTGAAGACCGGCAAGTACGCAAGATGGCCACCGCCACACCAACGAGGTG 1767
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGGCGCTGCAAGAAGATCGCCATCGAGAGCATCTGTGATCTGGGCAAG 1827
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540
QY 1828 ACCCCCAAGTTCGCTGCCCATCCAGAAAGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr 560
QY 1888 TGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCGCTGTGTGAAGCTG 1947
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1948 TGTGTACAGCTGGAGAAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGGACGGCGCC 2007
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2008 GCCAACCGCGAGACCAAGATCGCAAGCGCGCTACGTGACCCAGCCGCGGCGCGCAGAG 2067
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2068 ATCGTGAGCTGACCGCAGACCCACCAACAGAGACCGAGCTGCAGCGCATCCAGCTGGCC 2127
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
QY 2128 CTCACGACAGCGCGCAGCGAGGTGAACCTCGTCAACGACAGCCAGTACGCTGGGCATC 2187
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGGCCAGCCCAAGACGAGAGCGAGCTGGTGAACCCAGATCATCGAGCAGCTG 2247
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGGAGAGGTGTACCTGAGCTGGTGGTGGCCGCCACCAAGGGCATCGCGGCAAC 2307
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2308 GAGCAGATCGACAAGCTGGTGAAGAGGCATCCGCAAGGTGCTGTCTCTGGACGGCATC 2367
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db |||||
721 Asp 721
RESULT 14
AAP81861
ID AAP81861 standard; protein; 1002 AA.
XX
AC AAP81861;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 16-DEC-1990 (first entry)
XX
XX Sequence encoded by LAV MA L POL gene.
XX HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
XX Lymphadenopathy-associated virus.
XX PN W08707906-A.
XX
XX 30-DEC-1987.
XX
XX 22-JUN-1987; 87WO-RP000326.
XX
```

PR 23-JUN-1986; 85EP-00401380.  
 XX (INSP ) INST PASTEUR.  
 XX Alizon M, Sonigo P, Wainhobson S, Montagnier L, Alizon M;  
 XX Sonigo P, Wainhobson S, Montagnier L;  
 XX WPI; 1988-014396/02.  
 XX N-PSDB; AAN80437.  
 XX New variants of lymphadenopathy associated virus (LAV) - used for prodn.  
 XX of DNA, antigens and antibodies used in diagnosis of AIDS and pre-AIDS.  
 XX Claim 8; Fig 8A-8I; 72pp; English.  
 XX LAV EL I (AAN80436) and LAV MA L (AAN80437) were isolated from the  
 XX peripheral blood lymphocytes of patients. Different AIDS virus isolates  
 XX concerned are designated by 3 letters of the patients name. Stable probes  
 XX including the DNA sequences can be used for detection of the new LAV  
 XX viruses or related viruses or DNA proviruses in eg. biological samples.  
 XX The proteins or peptides can be used for detection of antibodies induced  
 XX in vivo and present in biological fluids. The DNA can also be used for  
 XX the expression of LAV viral antigens for the prodn. of a vaccine against  
 XX LAV. The polypeptides can also be used for the prodn. of antibodies for  
 XX the detection of proteins related to the LAV viruses, partic. for  
 XX diagnosis of AIDS or pre-AIDS. (Updated on 25-MAR-2003 to correct PI  
 XX field.) (Updated on 27-AUG-2003 to correct OS field.)  
 XX SQ Sequence 1002 AA;

Alignment Scores:  
 Pred. No.: 6.87e-199 Length: 1002  
 Score: 3532.50 Matches: 656  
 Percent Similarity: 96.53% Conservative: 39  
 Best Local Similarity: 91.11% Mismatches: 22  
 Query Match: 77.08% Indels: 3  
 DB: 1 Gaps: 2

US-09-610-313B-30 (1-2469) x AAP81861 (1-1002)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCGAGCCGCGAGTTCCCGAGGAGCAG 279  
 DB 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20  
 QY 280 AACCGCGCCAAACGCCACCGCGCGAGCTGCGAGTG---CGCGCGCAACAACCCCGC 336  
 DB 21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40  
 QY 337 AGCGAGCGCGCGCGAGCGCGAGGCG---ACCTGGAACCTTCCCGCGAGATCACCTG 390  
 DB 41 SerGluThrGlyAlaGluArgGlnGlyLeValSerPheSerPheProGlnThrLeu 60  
 QY 391 TGGCAGCGCGCGCGCGAGCATCAAGTGGCGCGCGAGATCAAGAGGCGCGCTGCTGAC 450  
 DB 61 TrpGlnArgProValThrValArgValGlyGlnLeuLysGluAlaLeuLeuAsp 80  
 QY 451 ACCGCGCGCGAGCAGCAGCTGCTGGAGGAGATGAGCCTCGCGCGCAAGTGAAGCCCAAG 510  
 DB 81 ThrGlyAlaAspThrValLeuGluGluLeuAsnLeuProGlyLysTrpLysProLys 100  
 QY 511 ATGATCGCGCGCGCGCGCTTCATCAGTGGCGCGAGTACGACCGAGATCCTGATCGAG 570  
 DB 101 MetileGlyGlyLeuGlyPheLeuLysValArgGlnTrpAspGlnLeuLeuLeuGlu 120  
 QY 571 ATCTGCGCGAGGAGCGCATCGCGCGCTGATCGCGCGCGCGCGCGCGCGCGCATCATC 630  
 DB 121 IleCysGlyLysAlaLeuGlyThrLeuValGlyProThrProValAsnLeuLeu 140  
 QY 631 GGCGCGCAACATGCTGACCGCGCTGGCTGACCTGACCTGACCTTCCCGCGCGCGCGCGCG 690  
 DB 141 GlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProLysProLysLeu 160  
 QY 691 ACCGTGCGCGTGAAGTGAAGCGCGCGCATGCGAGCGCGCGCGCGCGCGCGCGCGCTG 750

DB 161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180  
 QY 751 ACCGAGGAGAAGATCAAGCGCTGACCGCCATCTCGAGGAGATCGAGAGGAGGCGCAAG 810  
 DB 181 ThrGluGluLysLeuLysAlaLeuThrGluLeuLysAspMetGluLysGluGlyLys 200  
 QY 811 ATCAACCAAGATCGCGCGCGAGAACCCCTTACCAACACCCCGTGTTCGCCATCAAGAGAG 870  
 DB 201 IleLeuLysIleGlyProGluAsnProTrpAsnThrProValPheAlaIleLysLysLys 220  
 QY 871 GACAGCCACAGTGGCGCGAGCTGGTGGACTTCCGCGAGCTGACCAAGCGCGCACCCGAGC 930  
 DB 221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240  
 QY 931 TTCTGGAGGTGACGTGGCGCATCCCCACCCCGCGCGCTGAAAGAAAGAGAGCGGTG 990  
 DB 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
 QY 991 ACCGTGTGGACGTGGCGCGACCTTCTTACGCTGCGCGCTGGAGGAGCTTCCGCAAG 1050  
 DB 261 ThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArgLys 280  
 QY 1051 TACACCGCTTACCATCCCGCGCATCAACAGAGACCCCGCGCATCGGTACCGTACAGTAC 1110  
 DB 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300  
 QY 1111 AACGTGTGCGCGCGAGGTGGAAGCGAGCGCGCGCGCGCGCTTCCAGAGCAGATGACCAAG 1170  
 DB 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerMetThrLys 320  
 QY 1171 ATCTGGAGCGCTTCCCGCGCGCAACCCCGAGATCGTGATCTACAGTACATGAGCGAC 1230  
 DB 321 IleLeuGluProPheArgThrLysAsnProGluLeuIleValIleTyrGlnTyrMetAsp 340  
 QY 1231 CTGTACGTGGCGAGCGCTGGAGATCGCGCGAGCACCGCGCGCAAGATCGAGAGTGGCG 1290  
 DB 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuArg 360  
 QY 1291 AAGCACCTGTGCTGGCGCTTCCACCAACCCCGCGAGAGCAAGAGAGAGAGCGCGCG 1350  
 DB 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
 QY 1351 TTCTGTGGATGGCTACGAGCTGACCCCGCGAGTGGAGCGCGCTGCGAGCGCGCGAGCTG 1410  
 DB 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
 QY 1411 CCGAGAGAGAGAGCTGGACCGTGAAAGCATCCAGATCCAGAGCTGGTGGCGCAAGTGG 1470  
 DB 401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
 QY 1471 GCGAGCGAGATCTACCCCGCGCATCAAGTGGCGCGAGCTGTGCAAGCTGCTCGCGCGCG 1530  
 DB 421 AlaSerGlnIleLysProGlyLysValLysGlnLeuLysLysLeuLeuArgGlyAla 440  
 QY 1531 AAGCGCTGACCGAGATCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1590  
 DB 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuLeuProGluAsn 460  
 QY 1591 CGCGAGATCTGCGCGAGCG 1650  
 DB 461 ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
 QY 1651 GCGAGATCCAGAGAGCG 1710  
 DB 481 AlaGluIleGlnLysGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 500  
 QY 1711 AAGAACCTGAAGACCGCGCAAGTACGCCAAGATGCGCGCGCGCGCGCGCGCGCGCGAG 1770  
 DB 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
 QY 1771 CAGCTGACCGAGCGCGTGCAGAGATCGCATCGAGAGCGATCGTGATCTGGGCGCAAGAC 1830

Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
 QY 1831 CCACAGTTCGCTCCATCCAGAGGAGACCTGGGAGACCTGGGAGACCTACTGG 1890  
 Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTrp 560  
 QY 1891 CAGGCGACCTGGATCCCGGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGG 1950  
 Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580  
 QY 1951 TACCAGCTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTACCTGGAGCGCGGCC 2010  
 Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 QY 2011 AACCGGAGACCAAGATCGCAAGGCGGTAGTGAACACCGGCGGCGGCGAGATC 2070  
 Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
 QY 2071 GTGAGCTGACCGAGACCAACCAAGAACACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130  
 Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
 QY 2131 CAGGACAGCGCGAGCGAGGTGAACATCGTGACCGACGACGACGACGCTGGGCGATCATC 2190  
 Db 641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 QY 2191 CAGGCGGCGCGACAGCGGAGCGAGCTGGTGAACACGATCATCGAGCGCTGATC 2250  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680  
 QY 2251 AAGAAGGAGAGGTGTACCTGAGCTGGGTGCGCGCCGACCAAGGCGCATCGCGCGCAACGAG 2310  
 Db 681 GlnLysAspLysValLysLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnGlu 700  
 QY 2311 CAGATCGACAAAGTGTGTGACAAAGGCGCATCGCAAGTGCTGTCTCTGGAGCGCATCGAT 2370  
 Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 15  
 ABR55489  
 ID ABR55489 standard; protein; 1003 AA.  
 XX AC ABR55489;  
 XX DT 11-AUG-2003 (first entry)  
 XX DE Amino acid sequence of a HIV pol protein.  
 XX KW Polymorphic marker; host response; HIV; AIDS related virus; vaccine;  
 XX KW hepatitis related virus; HCV; HBV; drug resistance; pol.  
 XX OS Human immunodeficiency virus.  
 XX PN WO2003035097-A1.  
 XX PD 01-MAY-2003.  
 XX PF 23-OCT-2002; 2002WO-AU001450.  
 XX PR 23-OCT-2001; 2001AU-00008425.  
 XX PA (EPIP-) EPIPOT PTY LTD.  
 XX PI Mallal S;  
 XX DR WPI; 2003-449231/42.  
 XX PT Determining the influence of variation in host genes on the selection of  
 PT microorganisms with protein substitutions, comprises typing individuals  
 PT of a cohort infected with a microorganism for an intrinsic polymorphic  
 PT marker.  
 XX Claim 22; Page 88-89; 157pp; English.

XX CC The specification describes a method of determining the influence of  
 CC variation in host genes on selection of microorganisms with protein  
 CC substitutions. The method comprises typing all individuals of a  
 CC population of patients infected with a microorganism for at least one  
 CC selected intrinsic polymorphic marker involved in the host response to  
 CC the presence of the microorganism. The method is useful for examining  
 CC selective pressures confronting a wide range of organisms that exhibit  
 CC pathogenic traits in a host, such as bacteria, fungi, mycobacterium,  
 CC viruses and virus-like particles; for examining microorganisms that have  
 CC adapted to evolve rapidly, including HIV and AIDS related viruses and the  
 CC hepatitis related viruses such as HCV and HBV. The method is useful for  
 CC designing a vaccine to prevent or delay the emergence of drug resistance  
 CC in patients treated with a particular drug specific for a microorganism  
 CC where the drug affects the replication of the microorganism at the  
 CC nucleotide or amino acid level. The present sequence represents a HIV pol  
 CC protein, which is expected to provide optimal cytotoxic T lymphocyte  
 CC (CTL) induced therapeutic protection to the cohort examined in that study  
 XX SQ Sequence 1003 AA;  
 Alignment Scores:  
 Pred. No.: 7,35e-199 Length: 1003  
 Score: 3532.00 Matches: 658  
 Percent Similarity: 96.26% Conservative: 36  
 Best Local Similarity: 91.26% Mismatches: 23  
 Query Match: 77.07% Indels: 4  
 DB: 6 Gaps: 2  
 US-09-610-313B-30 (1-2469) x ABR55489 (1-1003)  
 QY 220 TTCTTCCGCGAGGACCTGGCTTCCCGGAGGCAAGCCGCGAGTTCCCGAGGAGCAG 279  
 Db 1 PhePheArgGlnAsnLeuAlaPheProGlnGlyValAlaArgGluPheSerSerGluGln 20  
 QY 280 AACCGGCGCAACACCGCCACGACCGCGAGCTCAGGTGCGCGCGAC-----AACCCC 333  
 Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGluAspAsnAsnSer 40  
 QY 334 CGCAGCGAGGCG 387  
 Db 41 ThrSerGlyAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60  
 QY 388 CTGTGGCAGCG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCG 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCG 567  
 Db 101 LysMetIleGlyIleGlyIleGlyIleGlyIleLysValArgGlnTyrAspGlnIleIle 120  
 QY 568 GAGATCTGCGGCAAGAGGCG 627  
 Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGGCGCGCAACATGCTGACCCAGCTGGGTGTCACCTGAACTTCCCATCAGCCCCCATC 687  
 Db 141 IleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACGCTGCG 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGAGGAGAGATCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCAACCAAGATCG 867

201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
Db  
868 AAGGACAGACCAAGTGGCGAAGCTGTGACTTCCCGAGCTGAACAAGCGCACCCAG 927  
Qy  
221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
Db  
928 GACTCTCTGGGAGGTGACGTGGGATCCCCACCCCGCGCGCTGAAGAAGAAGAGAGC 987  
Qy  
241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
Db  
988 GTGACCGTGTCTGGAGCTGGCGAGCCCTACTTCCAGCGTGGCCCTGGAGCAGACTTCCGC 1047  
Qy  
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
Db  
1048 AAGTACACCGCTTCCACCATCCCGAGCATCAACACAGAGACCCCGCGCATCCGCTACCCAG 1107  
Qy  
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
Db  
1108 TACACGTGTGCTCCCGAGGTGGAAGGGGAGCCCGAGCATCTCCAGAGCAGCATGACC 1167  
Qy  
301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
Db  
1168 AAGATCTGTGAGCCTTCCCGCGCGCGCAACCCCGAGATCGTGATCTACCACTACATGAC 1227  
Qy  
321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
Db  
1228 GACTGTGCTGAGGAGCAGCTCGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTG 1287  
Qy  
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
Db  
1288 CGAAGACCTGTGCTGGGTGGGTTCACACCCCGAGCAAGAGACCCAGAGAGAGCC 1347  
Qy  
361 ArgGlnHisLeuLeuLysTrpGlyPheThrProAspLysLysHisGlnLysGluPro 380  
Db  
1348 CCGTCTCTGTGTGAGTGTGAGCTGCAGCTGCACCCCGAGAGTGCAGCGCCCATCGAG 1407  
Qy  
381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
Db  
1408 CTGCCCCGAGAGAGAGCTGAGCCGTGAACACATCCAGAGCTGTGGGCAAGCTGAAC 1467  
Qy  
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
Db  
1468 TGGGCCAGCCGATCTACCCCGGATCAAGTGGCGCAGCTGTGCAAGCTGTCTGGCGGC 1527  
Qy  
421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuLysLysLeuLeuArgGly 440  
Db  
1528 GCCAAGGCCCTGACCGACATGCTGCCCTGACCCGAGGAGCGAGCTGGAGCTGGCGGAG 1587  
Qy  
441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460  
Db  
1588 AACCGCAGATCTCTGCGCAGCGCTGACGCGCTGTACTACGACCCCGAGAGGAGCTG 1647  
Qy  
461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480  
Db  
1648 GTGGCCGAGATCCAGAGAGGCGCCAGCAGCAGTGGACCTTACAGATCTACAGAGGCC 1707  
Qy  
481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500  
Db  
1708 TTCAAGAACCTGAGACCGCGAGTACGCCAAGTGGCCACCGCCGAGCAGCAGCAGCTG 1767  
Qy  
501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
Db  
1768 AAGCAGCTGACCGAGGCGGTGCAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827  
Qy  
521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
Db  
1828 ACCCCCAGTTCGCGCTGCCATCCAGAGAGAGACCTGGGAGACCTGGTGACCGACTAC 1887  
Qy  
541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560  
Db  
1888 TGGCAGGCCACCTGGATCCCGGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1947  
Qy  
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
Db

1948 TGGTACCAGCTGGAGAAAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCC 2007  
Qy  
581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
Db  
2008 GCCAACCCGAGAGACCAAGATCGGAAGCCCGGTACGTGACCGAGCCGGGGCGGAGAG 2067  
Qy  
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620  
Db  
2068 ATCTGTAGCTGACCGAGAGACCACCAACAGAGAGCCGAGCTGCAGGCCCATCCAGCTGGC 2127  
Qy  
621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
Db  
2128 CTGACGACACCGCGAGCGAGGTCAACATCTGTGACCGACACCGACCTGCGGCATC 2187  
Qy  
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
Db  
2188 ATCCAGGCCCGCGAG 2247  
Qy  
661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
Db  
2248 ATCAAG 2307  
Qy  
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
Db  
2308 GAGCAGATCCGACAGCTGTGTGAGCAAGGCGCATCCGCAAGGTGCTGTTCTTCTGGACGGCATC 2367  
Qy  
701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
Db  
2368 GAT 2370  
Qy  
721 Asp 721  
Db

Search completed: June 2, 2005, 04:41:43  
Job time : 256.242 secs



***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 11:33:41 ; Search time 167.833 Seconds  
(without alignments)  
11351.628 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 4573  
Sequence: 1 gtcagccaccatgcccga.....gggctagcaccggtgaattc 2463

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09610313/runat\_31052005\_151136\_151139/app\_query.fasta\_1.7893  
-DB-A\_Geneseq\_16Dec04 -QPMF=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -QPMF=score-pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US09610313 @CGN 1 1 557 @runat\_31052005\_151136\_151139 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -YGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Geneseq\_16Dec04:\*
- 1: Geneseq\_1980s:\*
- 2: Geneseq\_1990s:\*
- 3: Geneseq\_2000s:\*
- 4: Geneseq\_2001s:\*
- 5: Geneseq\_2002s:\*
- 6: Geneseq\_2003as:\*
- 7: Geneseq\_2003bs:\*
- 8: Geneseq\_2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3766.5	82.4	1435	8 ADS19486	Adp19486 Modified
2	3744.5	81.9	3183	8 ADP84803	Adp84803 HIV-1 hyb
3	3711.5	81.2	1457	8 ADN36406	Adn36406 Human pro
4	3620	79.2	1003	3 AAB69289	Aab69289 HIV-1 non
5	3611	79.0	999	3 AAB69286	Aab69286 HIV-1 non
6	3606	78.9	1005	3 AAB69287	Aab69287 HIV-1 non
7	3570	78.1	998	6 AAE37601	Aae37601 HIV-1 sub
8	3563	77.9	3025	4 AAB86169	Aab86169 HIV-1 sub
9	3521	77.0	854	5 AAM48949	Aam48949 HIV-1 sub
10	3517.5	76.9	1000	3 AAB69282	Aab69282 HIV-1 non

11	3499.5	76.5	1002	2 AAW72993	Aaw72993 HIV isola
12	3497	76.5	1003	6 AAO30963	Aao30963 HIV pol p
13	3497	76.5	1003	8 ADN36413	Adn36413 HIV prote
14	3494.5	76.4	1002	1 AAP81861	Aap81861 Sequence
15	3494	76.4	1003	6 ABR55489	Abr55489 Amino aci
16	3492	76.4	1003	1 AAP60420	Aap60420 Sequence
17	3492	76.4	1003	1 AAP70861	Aap70861 Sequence
18	3486	76.2	1003	3 AAY70602	Aay70602 Codon opt
19	3486	76.2	1003	3 AAY70601	Aay70601 Corrected
20	3484	76.2	1003	5 AAO19387	Aao19387 Lymphaden
21	3478.5	76.1	1003	2 AAR08060	Aar08060 HIV-1 pol
22	3475	76.0	1015	1 AAP60347	Aap60347 HTLV-III
23	3475	76.0	1015	2 AAR43867	Aar43867 HTLV-III
24	3475	76.0	1015	4 AAB85993	Aab85993 Amino aci
25	3472	75.9	1003	3 AAB10047	Aab10047 HIV-1 pol
26	3472	75.9	1003	3 AAY70600	Aay70600 Wild type
27	3471	75.9	1016	2 AAR08062	Aar08062 ACNEV-HIV
28	3468	75.8	1015	2 AAR43875	Aar43875 HTLV-III
29	3464	75.7	1016	2 AAR08063	Aar08063 HIV-1 pol
30	3461	75.7	2033	2 AAR08056	Aar08056 HIV-1 pol
31	3457	75.6	1016	2 AAB69284	Aab69284 HIV-1 pol
32	3455	75.6	1003	3 AAB08054	Aab08054 HIV-1 non
33	3454	75.5	2033	2 AAR08055	Aar08055 HIV-1 pol
34	3453	75.5	1003	8 ADP20076	Adp20076 Human imm
35	3449	75.4	739	5 AAU11874	Aau11874 HIV pol p
36	3449	75.4	1003	1 AAP61508	Aap61508 Sequence
37	3449	75.4	1003	2 AAR29705	Aar29705 pol gene
38	3449	75.4	1003	3 AAY77301	Aay77301 HIV-1 (AT
39	3449	75.4	1003	5 AAE35789	Aae35789 ARV-2 (9B
40	3449	75.4	1003	6 ABU63185	Abu63185 Pol prote
41	3449	75.4	1491	1 AAP91048	Aap91048 Transcrip
42	3446	75.4	1003	3 AAB69279	Aab69279 HIV-1 non
43	3444	75.3	1003	6 ABU57552	Abu57552 AIDS asso
44	3442	75.3	1012	6 ABU63325	Abu63325 Human lym
45	3439.5	75.2	1003	2 AAR08059	Aar08059 HIV-1 pol

ALIGNMENTS

RESULT 1

ADS19486  
ID ADS19486 standard; protein; 1435 AA.

AC ADS19486;

DT 18-NOV-2004 (first entry)

DE Modified HIV-1 GagPol protein.

KW Gene delivery; epithelial cell; respiratory system; mammal;  
KW lentiviral packaging system; expression vector; gagpol; gene expression;  
KW lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;  
KW cystic fibrosis; CF; deltaPD; transepithelial potential difference;  
KW gene therapy; HIV-1; mutant; mutain.

OS Human immunodeficiency virus 1.

OS Synthetic.

PN US2004037780-A1.

PD 26-FEB-2004.

XX 23-AUG-2002; 2002US-00226638.

XX 19-NOV-2001; 2001AU-00008942.

XX (PARS/) PARSONS D.

XX (ANSO/) ANSON D.

XX (LIMB/) LIMBERIS M.

XX (FULL/) FULLER M.

XX Parsons D, Anson D, Limberis M, Fuller M;

DR WPI; 2004-191004/18.  
DR N-PSDB; ADS19485.  
XX  
PT Delivering exogenous (e.g. CFTR) genes to epithelial cells in the  
PT respiratory tract using a penetration agent and a lentiviral packaging  
PT system, useful for treating e.g. cystic fibrosis.  
XX  
PS Disclosure; Fig 16; 45pp; English.  
XX  
CC The present invention relates to a method of delivering one or more  
CC exogenous genes for expression in an epithelial cell in the respiratory  
CC system of a mammal to give persistent expression of the gene in the  
CC epithelial cell, using a lentiviral packaging (expression vector) system.  
CC Also disclosed are (i) a recombinant lentiviral packaging system,  
CC comprising a first nucleic acid molecule including a gag gene sequence  
CC encoding a gag protein, and a second nucleic acid molecule including a  
CC gagpol gene sequence encoding a GagPol protein (the gagpol gene  
CC has degenerative nucleotide changes in the frame shift sequence AUUUUU  
CC to reduce the chance of a frame shift which switches expression of the  
CC Gagpol protein to the Gag protein in wild type lentivirus and the  
CC packaging system additionally comprises a lentiviral vector nucleic acid  
CC molecule which does not encode either the gag gene or the gagpol gene or  
CC both, and (ii) a recombinant nucleic acid molecule encoding a lentiviral  
CC gagpol gene having degenerative nucleotide changes in the frame shift  
CC sequence AUUUUU to reduce the chance of a frame shift which switches  
CC expression of the Gagpol protein to the Gag protein in the wild type  
CC lentivirus. The exogenous gene is expressed in sufficient numbers of  
CC cells and amounts to provide an ameliorating effect for a respiratory  
CC condition. The exogenous gene is cystic fibrosis transmembrane  
CC conductance regulator (CFTR) and the condition is cystic fibrosis  
CC ((CF)the expression of the CFTR gene is sufficient to provide a  
CC significant shift of a reduced deltaPD (transepithelial potential  
CC difference) back to normal levels in the mammal). The cell is non-  
CC terminally differentiated and is capable of differentiating into 2 or  
CC more cell classes e.g. ciliated cells, non-ciliated cells, secretory  
CC cells and basal cells. The exogenous gene may be enzymic. The gagpol gene  
CC sequence has additionally degenerate nucleotide substitutions which  
CC destabilise the hairpin structure associated with the frame shift event.  
CC The gag gene is a truncation of the wild type gagpol gene so that it can  
CC no longer be translated to form Gagpol. The lentivirus is HIV or HIV  
CC derived. The method and recombinant lentiviral packaging system may be  
CC used for delivering nucleic acids to epithelial cells in the respiratory  
CC system of a mammal. In particular they may be used to deliver the CFTR  
CC gene for the treatment of CF. It has been found that the delivery of a  
CC recombinant lentivirus carrying an exogenous gene to the respiratory  
CC system, following the delivery of a non-toxic amount of a penetration  
CC enhancer can provide persistence of expression of a gene product. The  
CC present sequence represents modified HIV-1 Gagpol protein.  
XX  
SQ Sequence 1435 AA;

Alignment Scores:  
Pred. No.: 2,87e-209 Length: 1435  
Score: 3766.50 Matches: 702  
Percent Similarity: 94.31% Conservative: 44  
Best Local Similarity: 88.75% Mismatches: 38  
Query Match: 82.36% Indels: 7  
DB: 8 Gaps: 4

US-09-610-313B-31 (1-2463) x ADS19486 (1-1435)

QY	13	ATGCGCGAGCCATGAGCCAGGCCACC--AGCGCAACATCTGTGATGCGAGCGCAAC	69
Db	363	LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn	382
QY	70	TTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGGAGGGCCCATGCGC	129
Db	383	PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla	402
QY	130	CGCAACTGCGCGCCCCCGCAGAGAGGGCTGTCTGGAAAGTGGCGCAAGGAGGGCCACCAG	189
Db	403	LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln	422
QY	190	ATGAGGAGTGCACCCAGCGCCAGGCCAACTTCTTCCGGAGGACCTGGCTTCCCCCAG	249
Db	423	MetLysAspCysThrGluArgGlnAlaAsnPheLeuArgGluAspLeuAlaPheProGln	442
QY	250	GGCAAGCCCGCGAGTTCCAGCAGAGAACCGCGCCAAACAGCCCCACCGACCGCGAG	309
Db	443	GlyLysAlaArgLysPheSerGluGlnThrArgAlaAsnSerProIleArgArgGlu	462
QY	310	CTGCAGGTG-----CGCGCGACACCCCGCGAGCGCGCGCGCGCGCGCGCGCG	363
Db	463	ArgGlnValTrpArgArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGly	482
QY	364	ACCTGT-----AACTTCCCCAGATCACCTGTGTGCAGCGCCCTGTGTGACATCAAG	417
Db	483	ThrValSerPheSerPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLys	502
QY	418	GTGGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCCGACACACCTGTGTGGAG	477
Db	503	IleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGlu	522
QY	478	GAGATGAGCTGCCCGCAAGTGGAGGCCAAGATGATCGCGCGCATCGCGCGCTTCATC	537
Db	523	GluMetAsnLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGlyPheIle	542
QY	538	AAGTGTGCGCATGACACCGATCTCTGATCGAGATCTGCGCGCAAGAGGCCATCGGCACC	597
Db	543	LysValArgGlnTyraAspGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThr	562
QY	598	GTGCTGATGCGCCCAACCCCGTGAACATCATCGCGCGCCGACATGTGACCCAGCTGGGC	657
Db	563	ValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGly	582
QY	658	TGCACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGAGCTGAGAGCCCGGC	717
Db	583	CysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGly	602
QY	718	ATGACGCGGCCCAAGCTGAAGCAGTGGCGCCCTCACCAGGAGAGATCAAGGCCCTGACC	777
Db	603	MetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuVal	622
QY	778	GCATCTGCGAGAGATGAGAGAGGAGGAGATCACCAGATCGCGCCCGCGAGACCCC	837
Db	623	GluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnPro	642
QY	838	TACAAACCCCGTGTCCCATCAAGAGAGAGGAGCAGCAGCAGTGGCGCAAGCTGGTG	897
Db	643	TyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuVal	662
QY	898	GACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTTGGGAGGTGAGCTGGGCATCCCC	957
Db	663	AspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIlePro	682
QY	958	CACCCCGCGCGCTGAAGAGAGAGAGCGTGCCTGATGAGCTGGCGCGAGCGCTAC	1017
Db	683	HisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyr	702
QY	1018	TTTCAGTGTCCCTGCGAGCGAGGACTTCCGCAAGTACACCGCTTCCATCCCGCAGCATC	1077
Db	703	PheSerValProLeuHisGluAspPheArgLysTrpThrAlaPheThrIleProSerIle	722
QY	1078	AACAAAGAGACCCCGGCATCTCCCTACCAAGTACAAGTGTGTGCTGCCCGAGGGCTGAAAGGC	1137
Db	723	AsnAsnGluThrProGlyThrArgTyrGlnTyrAsnValLeuProGlnGlyTyrLysGly	742
QY	1138	AGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTTGAGCCCTTCCGCGCCCGCAAC	1197
Db	743	SerProAlaIlePheGlnSerSerMetThrThrIleLeuGluProPheArgLysGlnAsn	762
QY	1198	CCCGAGATCGTGTATCTTACCAG-----GCCCTGTGTGTGGCGCAGCGCTGGAGATC	1251
Db	763	ProAsnLeuValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIle	782
QY	1252	GGCCAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGACCTGTGCTGCGCTGGGCTTCACC	1311

```
Db 783 GlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuAtrGrpGlyPheThr 802
Qy 1312 ACCCCGACAAAGACCAAGAGAGCCCTTCTGTGGATGGGCTACGAGCTGCAC 1371
Db 803 ThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHis 822
Qy 1372 CCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGTGCACCGTGAAC 1431
Db 823 ProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsn 842
Qy 1432 GACATCCAGAGCTGTGGGCAAGCTGAACCTGGCCAGCCAGCATCTACCCCGGCATCAAG 1491
Db 843 AspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys 862
Qy 1492 GTGGCCGAGCTGTGCAAGCTCTCGGGCGCCCAAGGCCCTGACCGCATCTGTCGCCCTG 1551
Db 863 ValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeu 882
Qy 1552 ACCGAGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGATCTCTGCGGAGCCCGTGCAC 1611
Db 883 ThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis 902
Qy 1612 GCGTGTACTACGACCCCGAGGAGCTGGCCGAGAACCCGCGAGATCTCTGCGGAGCCCGTGCAC 1671
Db 903 GlyValTyrTrpAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGly 922
Qy 1672 CAGTGGACCTACCAAGTCTTACGAGGCGCTTCAAGAACCTGAAGACCGGCAAGTACGACC 1731
Db 923 GlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAla 942
Qy 1732 AAGATGGCGACCGCCCAACCAACGAGCTGAGGAGCTGACCGAGCGCGCTGCAGAGATC 1791
Db 943 ArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIle 962
Qy 1792 GCCATGAGAGCATCGTATCTGGGGCAGACCCCGCAAGTTCGCCCTGCCCATCCAGAG 1851
Db 963 AlaThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLys 982
Qy 1852 GAGACCTGGGAGACCTGTGTGACCGACTACTGCGAGCGCCACTGATCCCGAGTGGGAG 1911
Db 983 GluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGlu 1002
Qy 1912 TTCGTGAACACCCCGCTGGTGAAGCTGTGGTACCAAGCTGGAGAGAGCCCATCATC 1971
Db 1003 PheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIle 1022
Qy 1972 GGCGCCGAGACCTTCTACGTGGACGGCGCCGCCCAACCGCGAGACCAAGATCGGCAAGGCC 2031
Db 1023 GlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAla 1042
Qy 2032 GGCTACGTGACCGACCGGGCGCGCAGAGATCTGAGCTGACCGAGACCAACCAACAG 2091
Db 1043 GlyTyrValThrAsnLysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGln 1062
Qy 2092 AAGACCGAGCTGACGAGCCATCCAGCTGGCCCTGACGAGCAGCGGAGCGGTGAACATC 2151
Db 1063 LysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIle 1082
Qy 2152 GTGACCGACCGACAGTACCGCTGGCATCATCCAGGCGCCAGCCGACGAGGAGGAGC 2211
Db 1083 ValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspArgSerGluSer 1102
Qy 2212 GAGCTGTGACAGATCATCGACGCTGATCAAGAGGAGGTGTACCTGAGCTGG 2271
Db 1103 GluLeuValSerGlnIleIleGlnLeuIleLysLysGlnLysValTyrLeuAlaTrp 1122
Qy 2272 GTGCCCGCCCAAGAGGCGATCGCGGGCAACGAGAGATCGCAAGCTGGTGCAGCAAGGC 2331
Db 1123 ValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGly 1142
Qy 2332 ATCCGCAAGGTGTCTTCTCGGACCGCATCATC 2364
```

```
Db 1143 IleArgLysValLeuPheLeuAspGlyIleAsp 1153
RESULT 2
ADP84803
ID ADP84803 standard; protein; 3183 AA.
XX ADP84803;
AC ADP84803;
XX 23-SEP-2004 (first entry)
DT HIV-1 hybrid isolate pNL4-3 protein.
DE eliciting; inducing; immune response; HIV; antigen; non-pathogenic;
KW vaccination; vaccine; HIV-1.
XX Human immunodeficiency virus 1.
XX WO2004056391-A1.
XX 08-JUL-2004.
XX 19-DEC-2003; 2003WO-AU001705.
XX 20-DEC-2002; 2002AU-00953556.
XX 17-SEP-2003; 2003AU-00905067.
XX (UTNE-) UNIV NEW SOUTH WALES.
XX Kent SJ, Purcell DF, Boyle DB, Ramsay A, Thomson S, Ramshaw IA;
XX WPI; 2004-500267/47.
XX N-PSDB; ADP84804.
XX Eliciting or inducing in a mammal an immune response against HIV-1
XX subtype AE, B or C by administering sequential doses of a recombinant
XX plasmid and viral vectors containing the nucleic acid molecules encoding
XX the HIV antigens.
XX Disclosure; Fig 12; 280pp; English.
XX The invention relates to a novel method for eliciting or inducing in a
XX mammal an immune response directed to a virus, preferably Human
XX immunodeficiency virus (HIV). The method comprises sequentially
XX administering to the mammal one or more sequential doses of a recombinant
XX plasmid vector or recombinant viral vector or its derivative, into which
XX the nucleic acid molecules encoding all, part or a modified form of two
XX or more antigens of the virus are incorporated, and one or more optimized
XX CpG motifs, where the antigens have been rendered substantially non-
XX pathogenic. The invention further comprises: a method for treating or
XX preventing viral infection in a mammal; a method of vaccinating a mammal
XX against a viral pathogen; a method of eliciting or inducing, in a mammal,
XX an immune response directed to HIV; a vaccine capable of inducing an
XX immune response directed to a virus comprising the recombinant plasmid
XX vector or recombinant viral vector or its functional derivative; a
XX nucleic acid construct or its functional derivative comprising the
XX plasmid vector or recombinant viral vector; a plasmid vector; a
XX recombinant viral vector; a pharmaceutical composition comprising the
XX nucleic acid, plasmid vector or recombinant viral vector constructs. The
XX method is useful in eliciting or inducing in a mammal an immune response
XX against HIV-1 subtype AE, B or C. This sequence represents the protein of
XX the HIV-1 hybrid isolate pNL4-3 DNA of the invention.
XX Sequence 3183 AA;
SQ
Alignment Scores: 6.39e-208 Length: 3183
Pred. No.: 3744.50 Matches: 707
Score: 87.43% Conservative: 44
Percent Similarity: 82.31% Mismatches: 33
Best Local Similarity: 81.88% Indels: 75
Query Match: 8 Gaps: 6
DB:
US-09-610-313B-31 (1-2463) x ADP84803 (1-3183)
```

QY 13 ATGCGGAGGCGCATGAGCCAGCCACCAGC---GCCAATCTCTGATGAGCGCCAGCAAC 69  
Db 363 LeuAlaGluAlaMetSerGlnValThrAsnProAlaThrMetIleGlnLysGlyAsn 382  
QY 70 TTCGAAGGCGCCCAAGCCATCATCAAGTGTCTTCAACTGCGGCAAGAGAGGCGCCATCGCC 129  
Db 383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402  
QY 130 CGCAACTGCGCGCGCCCGCCAGAGAGGGTCTGTGGAAGTGGCGCAAGAGAGGCGCCACAG 189  
Db 403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422  
QY 190 ATCAAGAGACTGCACCGAGCCAGGCCAAC----- 219  
Db 423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 442  
QY 220 -----TCTCTCCCGAGGAC 234  
Db 443 GlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPhePheArgGluAsp 462  
QY 235 CTGGCTTCCCGAGGCAAGGCCCGAGTTCCTCCAGCGAGCAGAACCGCGCCAAACAGC 294  
Db 463 LeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSer 482  
QY 294 ----- 294  
Db 483 ProProGluGluSerPheArgPheGlyGluGluThrThrThrProSerGlnLysGlnGlu 502  
QY 295 -----CCACACCGCGCGAGTTCGACGGT----- 318  
Db 503 ProIleAspLysGluLeuTyProLeuAlaProThrArgArgGluLeuGlnValTrpGly 522  
QY 319 CGCGGCGACAAACCCCGCAGCAGGCGCGCGCGAGCGCCAGGCGACCTCTGAAC----- 372  
Db 523 ArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSer 542  
QY 373 -----TCCCGCCAGATCACCTCTGG 393  
Db 543 SerLeuArgSerLeuPheGlySerAspProSerSerGlnPheProGlnIleThrLeuTrp 562  
QY 394 CAGCGCCCTCTGTGATCATCAAGTGGCGGCGCAGATCAAGAGAGGCGCTGTGTGACACC 453  
Db 563 GlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThr 582  
QY 454 GGCGCCGACCAACCGTGTGTGAGGAGATGAGCTGCCCGCGCAAGTGGAAAGCCCAAGATG 513  
Db 583 GlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysProLysMet 602  
QY 514 ATCGGGGCATCGCGCTTTCATCAAGTGGCGGCGCAGTACCAACAGATCCTGTGAGATC 573  
Db 603 IleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGluIle 622  
QY 574 TGGCGCAAGAGGCGCATCGGCACCGTGTGATCGGCCACCGCCCGCGTGAACATCATCGGC 633  
Db 623 CysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGly 642  
QY 634 CGCAACATGTCACCCAGCTGGGTGACCGCTGACCGCTGAACTTCCCATCAGCCCCCATCGAGACC 693  
Db 643 ArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThr 662  
QY 694 GTCCCGGTGAAGTGAAGCCCGCATGGACCGGCCCGCCAGGTGAAGCGTGGCCCGCTGACC 753  
Db 663 ValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThr 682  
QY 754 GAGGAGAAGATCAAGGCCCTTGACCGCCATCTCGGAGGAGATGAGAGAGGAGGCAAGATC 813  
Db 683 GluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIle 702  
QY 814 ACCAAGATCGGCCCGGAGAACCCCTTCAACACCCCGCTGTTCGCCATCAAGAGAGAGGAC 873  
Db 703 SerLysIleGlyProGluAsnProLysAsnThrProValPheAlaIleLysLysLysAsp 722

QY 874 AGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTC 933  
Db 723 SerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPhe 742  
QY 934 TGGAGGTGCAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAAGAGAGCGTGACC 993  
Db 743 TrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSerValThr 762  
QY 994 GTCTCTGAGCGTGGGCGACGCTTACTTCCAGGTGCTCCCTGGACGAGGATTCCTCCAGTAC 1053  
Db 763 ValLeuAspValGlyAspAlaTrpPheSerValProLeuAspLysAspPheArgLysTrp 782  
QY 1054 ACCGCTTCAACATCCCGAGCAACAACAGACACCCCGGATCCGCTACCACTAGTCAAC 1113  
Db 783 ThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGlnTrpAsn 802  
QY 1114 GTCTCTCCCGAGGCTGGAGGCGACCCCGAGCATCTTCAGACGACGATGACCAAGATC 1173  
Db 803 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThrLysIle 822  
QY 1174 CTGGAGCCTTCCCGCGCCCGCAACCCCGAGATCGTGTATCTACCAG-----GCCCCCTG 1227  
Db 823 LeuGluProPheArgLysGlnAsnProAspIleValIleTrpGlnTrpMetAspAspLeu 842  
QY 1228 TAGCTGGGCGAGCGACCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGAGCTGCGCAAG 1287  
Db 843 TyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuLeuArgGln 862  
QY 1288 CACTCTCTCGCTGGGCTTCCACCCCGCACAGAAGCACCAAGAGAGGCCCTCTTC 1347  
Db 863 HisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPhe 882  
QY 1348 CTGTGATGGCTACAGCTGCGACCCCGCAGAGTGCAGCGTGCAGCCCATCGAGCTGCC 1407  
Db 883 LeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleValLeuPro 902  
QY 1408 GAGAAAGAGAGCTGGAGCGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACCTGGGCC 1467  
Db 903 GluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAla 922  
QY 1468 AGCCAGATCTACCCCGGCATCAAGTGGCGCCAGCTGTGCAAGCTGTGCGCGCGCAAG 1527  
Db 923 SerGlnIleTyAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLys 942  
QY 1528 GCCTGACCGACATCTGCGCCCTGACCGAGAGGCGCGAGCTGCGCGGAGACCGC 1587  
Db 943 AlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArg 962  
QY 1588 GAGATCTGCGGAGCGCGTGCACCGCGTGTACTAGACCCCGAGCAAGGACCTGGTGGCC 1647  
Db 963 GluIleLeuLysGluProValHisGlyValTyTrpAspProSerLysAspLeuIleAla 982  
QY 1648 GAGATCCAGAAGCAGGCGCCACGACGAGTGGACCTTACCAGATCTTACCAGGAGCCCTTCAAG 1707  
Db 983 GluIleGlnLysGlnGlyGlnTrpThrTyTrpGlnIleTyTrpGlnProPheLys 1002  
QY 1708 AACCTGAAGACCGGCAAGTACCGCAAGATGCGCACCGCCCGCACCAACGACGAGTGAAGCAG 1767  
Db 1003 AsnLeuLysThrGlyLysTyAlaArgMetLysGlyAlaHisThrAsnAspValLysGln 1022  
QY 1768 CTGACCGAGCGCTGCAGAAAGTCCGCATGCGAGAGATCGGTGATCTGGGGCAAGACCC 1827  
Db 1023 LeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPro 1042  
QY 1828 AAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTTGGTGGACCGACTACTGGCAG 1887  
Db 1043 LysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTrpGln 1062  
QY 1888 GCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGGTAC 1947  
Db 1063 AlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTy 1082  
QY 1948 CAGCTGGAGAAGGAGGCCCATCATCGCGCGCGAGACCTTCTTACGTGGACGCGCGCCCAAC 2007

Db 1083 GlnLeuGluLysGluProIleGlyAlaGluThrPheTyrValAspGlyAlaAlaAsn 1102  
 Qy 2008 CGCGAGACCAAGATCGGCAAGCGCGTACGTGACCGACCGGGCGCGACAGATCGTG 2067  
 Db 1103 ArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspAsgGlyArgGlnLysValVal 1122  
 Qy 2068 AGCTGACCGAGACCAACCAAGACAGACCGAGCTGAGCGCATCCAGTGGCCCTGAG 2127  
 Db 1123 ProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142  
 Qy 2128 GACAGCGCAGCGAGGTGAACATCGTGACCGACAGCAGCAGTACGCGCTGGGCATCATCCAG 2187  
 Db 1143 AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleLeuGln 1162  
 Qy 2188 GCCCAGCGCCGACAAAGAGCGAGCGAGCTGGTGAACACAGATCATCGACAGCTGATCAAG 2247  
 Db 1163 AlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuLys 1182  
 Qy 2248 AAGGAGAAGGTGTACGTAGCTGGGTGCCGCCCAAGAGGCGATCGCGGCAACAGAGCAG 2307  
 Db 1183 LysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysGlnGluGln 1202  
 Qy 2308 ATCCGACAGCTGTGAGCAAGGGCATCCGCAAGGTGTCTTCTGGAGCGGCATCGAT 2364  
 Db 1203 ValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1221  
 RESULT 3  
 ADN36406  
 ID ADN36406 standard; protein, 1457 AA.  
 AC ADN36406;  
 XX  
 DT 15-JUL-2004 (first entry)  
 XX  
 DE Human protein for anti-HIV vaccine.  
 XX  
 KW anti-HIV; vaccine; HIV; promoter; viral particle; immunization.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02004035006-A2.  
 XX  
 PD 29-APR-2004.  
 XX  
 PF 17-OCT-2003; 2003WO-US033112.  
 XX  
 PR 18-OCT-2002; 2002US-0419465P.  
 XX  
 PA (AARO-) AARON DIAMOND AIDS RES CENT.  
 XX  
 PI Huang Y, Ho DD, Chen Z;  
 XX  
 DR WPI; 2004-348328/32.  
 XX  
 DR N-PSDB; ADN36405.  
 XX  
 PT Nucleic acid vector comprising at least one HIV sequence operably linked  
 PT to a promoter and encoding a protein that does not assemble into viral  
 PT particles, useful in immunizing a subject against HIV infection.  
 XX  
 PS Disclosure; SEQ ID NO 20; 166pp; English.  
 CC  
 CC The invention relates to a nucleic acid vector comprising at least one  
 CC HIV sequence operably linked to a promoter and encoding a protein that  
 CC does not assemble into viral particles. The nucleic acid vector is useful  
 CC in immunizing a subject against HIV infection. This sequence corresponds  
 CC to a peptide used in the invention.  
 XX  
 SQ Sequence 1457 AA;  
 Alignment Scores:  
 Pred. No.: 4,34e-206 Length: 1457  
 Score: 3711.50 Matches: 706

Percent Similarity: 91.86% Conservative: 28  
 Best Local Similarity: 88.36% Mismatches: 45  
 Query Match: 81.16% Indels: 21  
 DB: 8 Gaps: 4  
 US-09-610-313B-31 (1-2463) x ADN36406 (1-1457)  
 Qy 13 ATGCGCGAGCGCATGAGCGGCAACAGCGCCCAACATCTGTATGCGAGCGGCAACATCTTC 72  
 Db 383 LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnPhe 401  
 Qy 73 AAGGGCCCCAAGCCATCATCAAGTGTCTTCACTGCGGCAAGGAGGGGCCACATCCCGCGC 132  
 Db 402 LysGlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 421  
 Qy 133 AACTGCGCGCGCCCGCAAGAGGGTGTCTGGAAGTCTGGCAAGAGGGGCCACCATGATG 192  
 Db 422 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 441  
 Qy 193 AAGGACTGACCGAGCGCGCAGGCAACATCTTCTCGCGAGAGACCTGGCTTCCCCCAGGGC 252  
 Db 442 LysAspCysThrGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProProHisLysG1 461  
 Qy 253 AAGGCGCGCGAGTTCGCCGAGCAGACAAACCGCCCAACAGCCCGGCGAGCGCGAGCTG 312  
 Db 461 YArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPh 481  
 Qy 313 CAGGTGCGCGCGAGCAACCGCGCGAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCG 372  
 Db 481 eGlyPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuTh 501  
 Qy 373 TTC-----CCCCAGATCACCCCTGTGCA 395  
 Db 501 rSerLeuLysSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpG1 521  
 Qy 396 GCGCGCGCGTGTGAGCATCAAGTGGGCGGCGCATCAAGGAGGCGCGCGCGCGCGCGCG 455  
 Db 521 nArgProLeuValSerIleArgValGlyGlyIleLysGluAlaLeuLeu----- 538  
 Qy 456 CCGCGAGCACACCGTGTGAGAGAGATGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 515  
 Db 539 ----AspAspThrValLeuGluValAsnLeuProGlyLysTrpLysProLysMetI1 557  
 Qy 516 CCGCGCGCATCGCGCTTCATCAAGGTGCGCGCATGACAGCAGATCTGTATGAGATCTG 575  
 Db 557 eGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleProLysGluIleCy 577  
 Qy 576 CGGCAAGAGCGCATCG 635  
 Db 577 sGlyLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597  
 Qy 636 CAACATGCTGACCGAGTGGGCTGCACCTGAACTTCCCATCATGAGCGCGCGCGCGCGCG 695  
 Db 597 gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProLysGluThrI1 617  
 Qy 696 GCGCGTGAAGTGAAGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755  
 Db 617 eProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrG1 637  
 Qy 756 GGAGAAGATCAAGCG 815  
 Db 637 uGluLysIleLysAlaLeuThrAlaIleCysAspGluMetGluLysGluGlyLysIleTh 657  
 Qy 816 CAAGATCGCGCGCGCGAGAACCCCTCAACACCGCGCGTGTCCCGCATCAAGAGAAGAGCAG 875  
 Db 657 rLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSe 677  
 Qy 876 CACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAGCGCGCGCGCGCGCGCG 935  
 Db 677 rThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTr 697  
 Qy 936 GGAGGTGCGAGTGGGCGATCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 995



697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrVa 717  
 996 GCTGGAGTGGGAGCCCTACTTCCAGCGGCCCTCGAGAGAGACTTCCGCAAGTACAC 1055  
 717 lLeuAspValGlyAspAlaTyrPheSerValProLeuTyrGluAspPheArgLysTyrTh 737  
 1056 CGCTTTCCACCATCCAGCATCAACACAGAGACCCCGGCATCCGTTACCAGTACCAACGT 1115  
 737 rAlaPheThrIleProSerIleAsnGlnThrProGlyIleArgTyrGlnTyrAsnVa 757  
 1116 GCTGCCAGGCTGGAAGCGCCAGCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175  
 757 lLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777  
 1176 GAGCCCTTCCGCGCCGCAACCCCGAGATCGTATCTACCAG-----GCCCCCTGTA 1229  
 777 uGluProPheArgAlaGlnAsnProGluIleValIleTyrGlnTyrGlyAspAspLeuTy 797  
 1230 CPTGGGCGAGGACTCGAGATCGGCAGACCCCGCCAGCATCGAGGAGCTGCGCAAGCA 1289  
 797 rValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHi 817  
 1290 CTTCTGCTGGGCTTACCAACCCCGCAAGAACAGACCAAGAGAGCCCTTCTTCT 1349  
 817 sLeuLeuLysTyrGlyPheThrThrProAspLysLysHisGlnLysGluProPheLe 837  
 1350 GTGGATGGCTAGAGCTGACCCCGACAGTGGCCGTCGTCAGCCCATCGAGCTGCCGA 1409  
 837 uTrpMetGlyTyrGluLeuHisProAspLysTyrThrValGlnProIleGlnLeuProGl 857  
 1410 GAAGGAGAGCTGGACCGTGAACACATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCCAG 1469  
 857 uLysAspSerTyrThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877  
 1470 CCAGATCTACCCCGCATCAAGTGGCCAGCTGTGCAAGCTCTGCGCGCCCAAGGC 1529  
 877 rGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897  
 1530 CTGACCGGACATCGTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGAGACCGCGA 1589  
 897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGl 917  
 1590 GATCTCGCGGAGCCGTCGACCGCGTGTACTACGACCCCGCAAGGACCTGTGTGGCCGA 1649  
 917 uIleLeuLysGluProValHisGlyAlaTyrTyrAspProSerLysAspLeuIleAlaGl 937  
 1650 GATCCAGAACGAGCGGCACACAGTGGACCTTACAGATCTACAGAGAGCCCTTCAAGAA 1709  
 937 uIleGlnLysGlnGlyAspGlnTyrThrTyrGlnIleTyrGlnGluProPheLysAs 957  
 1710 CTGAGACCGGCAAGTACCGCAAGATCGCACCGGCCACACCAACGACGTGAGCAGCT 1769  
 957 nLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977  
 1770 GACCGAGCGCTGCAGAAAGATCGCATCGTAGAGATCGTATCTGGGGCAAGACCCCA 1829  
 977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTyrGlyLysIleProLy 997  
 1830 GTTCCGCTCCCATCCAGAGAGACCTCGGAGACTGTGGACCGCACTACTGGCAGGC 1889  
 997 sPheArgLeuProIleProLysGluThrTyrGluThrArgTyrThrAlaTyrTrpGlnAl 1017  
 1890 CACTGGATCCCGAGTGGAGTTCGTGACACCCCCCTCGTGAAGCTGTGTACCA 1949  
 1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTyrTrpGl 1037  
 1950 GCTGGAGAAGGAGCCCATCATCTACGTCGAGACCTTCTAGCTGGAACCGCCGCCAACCG 2009  
 1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTyrValAspGlyAlaAlaAsnAr 1057  
 2010 CGAGACCAAGATCGCAAGCGCCGCTACTAGTACCGACCGCGCGCGGAGAGATCGTGAG 2069  
 1057 gGluThrLysMetGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSe 1077

2070 CCTGACCGAGACCACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGGA 2129  
 1077 rLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleCysLeuAlaLeuGlnAs 1097  
 2130 CAGCGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGC 2189  
 1097 pSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAl 1117  
 2190 CAGCGCCGACAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAA 2249  
 1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLy 1137  
 2250 GGAAGAGGTGTACTCTGAGCTGGGTGCCCGCCCAAGGGCATCGCGGCAACGAGCAGAT 2309  
 1137 sGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyCysGlnGluGlnVa 1157  
 2310 CGACAAGCTGGTGAGCAAGGCGCATCGCAAGGTGTCTTCTCGACGCGCATCGAT 2364  
 1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175  
 RESULT 4  
 AAB69289  
 ID AAB69289 standard; protein; 1003 AA.  
 XX AAB69289;  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 DE HIV-1 non-subtype B clone 94IN476-104 pol protein.  
 XX  
 KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 PN WC200026416-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 25-OCT-1999; 99WO-US024837.  
 XX  
 PR 02-NOV-1998; 98US-00184418.  
 XX  
 XX (UABR-) UAB RES FOUND.  
 XX  
 XX Hahn BH, Shaw GM, Gao F;  
 XX WPI; 2000-365651/31.  
 XX  
 XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.  
 XX  
 PS Claim 41; Fig 15; 131pp; English.  
 XX  
 CC The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1003 AA;  
 Alignment Scores:  
 Pred. No.: 7,76e-201 Length: 1003  
 Score: 3620.00 Matches: 677  
 Percent Similarity: 98.05% Conservative: 26  
 Best Local Similarity: 94.42% Mismatches: 12

Query Match:	79.16%	Indels:	2
DB:	3	Gaps:	1
US-09-610-313B-31 (1-2463) x AAB69289 (1-1003)			
QY	220	TTCTTCGCGAGGACCTTCCCTCCAGGCAAGCCCGCGAGTTCCCGAGGAGCAG	279
DB	1	PhePheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerLysGln	20
QY	280	AACCGCGCCAAACAGCCACAGCCGCGAGCTGCGAGTGCAGGCGACAAACCCCGCAGC	339
DB	21	AlaArgAlaAsnSerProThrSerArgGluLeuGlnValGlnGlyAspAsnProArgSer	40
QY	340	GAGCGCGCGCGAGCGCGAGGCGACCTGAACTTCCCGAGATCACCTGTGCGAGCGC	399
DB	41	GluAlaGlyValGluArgGlnGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg	60
QY	400	CCCTGTGTGAGCATCAAGGTGGCGCGCAGATCAAGGAGGCGCTGCGACACCGCGCC	459
DB	61	ProLeuValSerIleLysValGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla	80
QY	460	GACGACACCGTGTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATATCGGC	519
DB	81	AspAspThrValLeuGluGluIleAlaLeuProGlyArgTrpLysProLysMetIleGly	100
QY	520	GGCATCGCGCGCTTCATCAAGTGGCGGAGTACGACGAGTCCGTGATCGAGATCTCGGC	579
DB	101	GlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGly	120
QY	580	AAGAAGCCATCGGACCGTCTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAAC	639
DB	121	LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsp	140
QY	640	ATGTGACCCAGCTGGGTGACCCCTGACCTTCCCATCTCCAGCCCATCGAGACCGTCCC	699
DB	141	MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro	160
QY	700	GTGAAGTGAAGCCCGGATGAGCGGCCCAAGGTGAAGCGATGGCCCTGACCGAGGAG	759
DB	161	ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu	180
QY	760	AAGTCAAGCCCTGACCGCCATCTCGAGGAGTGGAGAGGAGGCGCAATCACCAAG	819
DB	181	LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys	200
QY	820	ATCGGCGCGAGAACCCCTACAACACCCCGTGTTCGCATCAAGAAGAGGAGCAGCAC	879
DB	201	IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysArgLysAspSerThr	220
QY	880	AAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCGGGAG	939
DB	221	LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu	240
QY	940	GTGAGCTGGGCATCCCGCCACCCCGCGGCTGAAGAAGAAGAGAGCGTGACCGTCTG	999
DB	241	ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu	260
QY	1000	GAGTGGCGGACGCTACTTTCAGGTGCGCCCTGACGAGGACTTCCGCAAGTACACCGCC	1059
DB	261	AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheGlyLysTyrThrAla	280
QY	1060	TTCAACATCCCGAGCATCAACAGAGACCCCGCGCATCCGCTACCAAGTACAACGTCTG	1119
DB	281	PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu	300
QY	1120	CCCGAGGCTGGAGGCGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAG	1179
DB	301	ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu	320
QY	1180	CCCTTCGCGCGCGCAACCCGAGATGTGATCTACCGAG-----GCCCGCTGTACGTG	1233
DB	321	ProPheArgAlaArgAsnProLysIleValIleTyrGlnTyrMetAspAspLeuTyrVal	340

QY	1234	GGCAGCGACCTGGAGATCGCGCAGCAGCGCCCAAGATCGAGGAGCTGCGCAAGCACCTG	1293
DB	341	GlySerAspLeuGluIleGlyHisArgAlaLysIleGluGluLeuArgAlaHisLeu	360
QY	1294	CTGCGCTGGGCTTCACACACCCCGCAAGAAGCACCAAGAAGGAGCCCTTCCTGTGG	1353
DB	361	LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp	380
QY	1354	ATGGGCTACGAGTGCACCCCGCAAGTGGACCTGCGAGCCCATCGAGTCCCGAGAG	1413
DB	381	MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleLysLeuProGluLys	400
QY	1414	GAGAGCTGGACCGTGAAGACATCCAGAGCTGGTGGCAGCTGAACCTGGCCAGCCAG	1473
DB	401	AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln	420
QY	1474	ATCTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGCCCTG	1533
DB	421	IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgLysAlaLysAlaLeu	440
QY	1534	ACCGACATCGTGGCCCTGACCGAGGCGCGAGCTGGAGTGGCGGAGAACCCGCGAGATC	1593
DB	441	ThrAspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIle	460
QY	1594	CTGCGCGAGCCGTGCGCGCGCTGTACTACACCCCGCAGCAGGAGCTGGTGGCCGAGATC	1653
DB	461	LeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIle	480
QY	1654	CAGAAGCAGCGCCACACAGTGGACCTTACCAGATCTACCAGAGGAGCCCTTCAAGAACCTG	1713
DB	481	GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu	500
QY	1714	AAGACCGGCAAGTACGCCAAGATGCGCACCGCCCGCACCAACAGCTGAAGCAGCTGACC	1773
DB	501	LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr	520
QY	1774	GAGCGCTGCAAGATCGCCATGGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTC	1833
DB	521	GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp***LysThrProLysPhe	540
QY	1834	CGCTGCGCCATCCAGAAGGAGACCTGGAGACCTGGTGGACCGACTACTGGCAGCGCAC	1893
DB	541	ArgLeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAlaThr	560
QY	1894	TGATATCCCGAGTGGGAGTTCTGTGAACACCCCGCTGGTGAAGCTGTGTACAGCTG	1953
DB	561	TrpIleProAspTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu	580
QY	1954	GAGAGGAGCCATCATCGCGCGAGACCTTCTACGTGAGCGCGCGCCCAACCGCGAG	2013
DB	581	GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu	600
QY	2014	ACCAAGATCGCGAGCGCGCTAGCTACCGACCGCGCGCGGCGGAGAGATCGTGAGCTG	2073
DB	601	ThrLysValGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeu	620
QY	2074	ACCGAGACCCCAACCAAGAGACCGAGCTGCGAGGCCATCCAGCTGGCGCTCGAGGACAGC	2133
DB	621	ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer	640
QY	2134	GGCAGCAGGTGAACATCTGTGACCGACGACCGCGCTACCGCTGGGCATCATCCAGGCCAG	2193
DB	641	GlyThrGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln	660
QY	2194	CCGACAAAGAGCAGGAGCGAGCTGGTCAACACAGATCATCCAGCAGCTGATCAAGAGGAG	2253
DB	661	ProAspLysSerGluSerGluLeuValAsnGlnIleIleGlnGlnLeuIleAsnLysGlu	680
QY	2254	AAGTGTACCTGAGCTGGTGGTGGCGCCCGCAAGGGCATCGCGCGCAACGAGCATCGAC	2313
DB	681	ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnGluGlnValAsp	700
QY	2314	AAGCTGTGAGCAAGGCGCATCCCGCAAGGTGCTGTCTCTGGACGCGCATCGAT	2364



Db 421 IletyAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440  
Qy 1534 ACCGACATCGTCCCTGACGAGAGAGCCGAGCTGGAGCTGGCGGAGAACCGGAGATC 1593  
Db 441 ThrAspIleValProLeuThrGluGluAlaGluLeuAlaGluAsnLysGluIle 460  
Qy 1594 CTGCGGAGCCGCGGACGGGTGCTACTACACCCAGCAGGAGCTGGTGGCGGAGATC 1653  
Db 461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480  
Qy 1654 CAGAAGCAGGCGCCACGACCTGACCTACACAGATCTACACAGGAGCCCTCAAGAACCTG 1713  
Db 481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500  
Qy 1714 AAGACCGGCAAGTACGCAAGATCGCACCGCCACACACGACGTAAGACGCTGACC 1773  
Db 501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520  
Qy 1774 GAGCCCTGCAGAGATCGCCATCGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTC 1833  
Db 521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTyrGlyLysIleProLysPhe 540  
Qy 1834 CGCTGCGCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTACTGGCAGGCCACC 1893  
Db 541 ArgLeuProLysGlnLysGluThrTyrGluThrTyrTrpThrAspTyrTrpGlnAlaThr 560  
Qy 1894 TGGATCCCGAGTGGGAGTTCTGTAACACACCCCGCCCTGGTGAAGCTGTGTACAGCTG 1953  
Db 561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTyrTyrGlnLeu 580  
Qy 1954 GAGAGGAGCCATCATCGGCGCGGAGACCTTCTACGTGAGCGGCGGCCCAACCGCGAG 2013  
Db 581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600  
Qy 2014 ACCAAGATCCGACGAGCGGCTAGTACGACGAGCGGCGGCGGAGAGATCGTGAGCTG 2073  
Db 601 ThrLysLeuGlyLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620  
Qy 2074 ACCGAGACCAACCAAGAGACCGAGCTGCAGGCCCATCCAGCTGGCGCCCTGCAGGACAGC 2133  
Db 621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640  
Qy 2134 GGCAGCAGGTGAATCGTGACGACGAGCAGCTGACGCGCTGGCGATCATCCAGGCCAG 2193  
Db 641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660  
Qy 2194 CCGACAAAGCAGGAGCGAGCTGTGAACACGATCATCCAGAGCTGATCAAGAAGGAG 2253  
Db 661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 680  
Qy 2254 AAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATCGGCGGCAACGAGCAGATCGAC 2313  
Db 681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnValAsp 700  
Qy 2314 AAGCTGTGAGCAAGGCGATCCGCAAGGTGTCTTCTGACCGGCATCGAT 2364  
Db 701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 6  
AAB69287  
ID AAB69287 standard; protein; 1005 AA.  
XX  
AC AAB69287;  
XX  
DT 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX  
DE HIV-1 non-subtype B clone 962W751-3 pol protein.  
XX  
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
KW vif; vpr; tat; rev; nef; vaccine.  
XX  
OS Human immunodeficiency virus 1.

XX WO200026416-A1.  
PN 11-MAY-2000.  
XX  
PF 25-OCT-1999; 99WO-US024837.  
XX  
PR 02-NOV-1998; 98US-00184418.  
XX  
PA (UABR-) UAB RES FOUND.  
XX  
PI Hahn BH, Shaw GM, Gao F;  
XX WPI; 2000-365651/31.  
DR  
XX  
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.  
XX  
PS Claim 41; Fig 15; 131pp; English.  
XX  
CC The present in invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 1005 AA;  
Alignment Scores:  
Pred. No.: 5e-200 Length: 1005  
Score: 3605.00 Matches: 681  
Percent Similarity: 96.68% Conservative: 18  
Best Local Similarity: 94.19% Mismatches: 16  
Query Match: 78.85% Indels: 8  
DB: 3 Gaps: 2  
US-09-610-313B-31 (1-2463) x AAB69287 (1-1005)  
Qy 220 TTCTTCGCGAGGACCTGGCGCTTCCCGGAGGCAAGGCCCGGAGTTCCCGAGGAGCAG 279  
Db 1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyGluLeuProSerGluGln 20  
Qy 280 AACCGCGCC-----AACAGCCCCACCGCCGCGAGCTCGAGTGGCGC 321  
Db 21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40  
Qy 322 GCGACAAACCCCGCGAGCG 381  
Db 41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60  
Qy 382 ATCACTGTGGAGCGCCCGCTGGTGGAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCG 441  
Db 61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIleLysGluAla 80  
Qy 442 CTGCTGGACACCG 501  
Db 81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrp 100  
Qy 502 AAGCCCAAGATGATCGGCGCGCATCGGCGCGCTTCAATCAAGGTGGCGCGCGCGCGCG 561  
Db 101 LysProLysMetIleGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIle 120  
Qy 562 CTGATCGAGATCTCGCGCAAGAGCGCATCGGACCGCTGTGTGTGTGTGTGTGTGTGT 621  
Db 121 LeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVal 140  
Qy 622 AACATCATCGCGCGCAACATGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 681  
Db 141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160



DR WPI: 2003-430497/40.  
XX N-PSDB: AAD29258.  
PT New molecules comprising HIV-1 subtype isolate regulatory/accessory  
PT genes, useful for manufacturing a vaccine for treating or preventing HIV  
PT infection.  
XX  
PS Disclosure: Page 94-97; 97pp; English.  
XX  
CC The invention relates to molecules comprising HIV-1 subtype isolate  
CC regulatory/accessory genes (tat, nef and rev genes) and modifications and  
CC derivatives thereof. The invention also provides proteins encoded by such  
CC genes. Sequences of the invention are useful for manufacturing vaccines  
CC for treating or preventing human immunodeficiency virus (HIV) infections.  
CC They are also useful in gene therapy. The present sequence is HIV-1  
CC subtype C isolate Du151 reverse transcriptase (RT) protein. (Updated on  
CC 23-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 998 AA;  
Alignment Scores:  
Pred. No.: 6.01e-198 Length: 998  
Score: 3570.00 Matches: 666  
Percent Similarity: 97.07% Conservative: 29  
Best Local Similarity: 93.02% Mismatches: 19  
Query Match: 78.07% Indels: 2  
DB: 6 Gaps: 1  
US-09-610-313B-31 (1-2463) x AAE37601 (1-998)  
QY 223 TTCCGGCAGGACCTGGCTTCCCTCCAGGGCAGGCGCGGAGTTCCTCCAGGAGCAGAAC 282  
DB 1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20  
QY 283 CGGCGCAACAGCCACACAGCGGAGCTGCAGGTGGCGCGACAAACCCCGCAGCGAG 342  
DB 21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnProArgSerGlu 40  
QY 343 GCCGGCGCGAGCGCCAGGCGACCTGAACCTCCCGCAGATCACCTGTGGCAGCGCCC 402  
DB 41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnThrLeuTrpGlnArgPro 60  
QY 403 CTGGTGAAGTCAAGTGGGGCGGCACATCAGAGAGCCCTGTGGACACCGGCGCGAC 462  
DB 61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80  
QY 463 GACACCGTGTGGAGGAGATGAGCTCCCGGCAAGTGAAGCCCAAGATGATCGGGCG 522  
DB 81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100  
QY 523 ATCGGCGGCTTCATCAAGGTGGCGCAGTACGACACAGATCCTGATCGAGATCTGCGGCAAG 582  
DB 101 IleGlyGlyPheIleLysValArgGlnThrAspGlnLeuIleGluIleCysGlyLys 120  
QY 583 AAGGCCATCGGCGCGCTGTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATG 642  
DB 121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140  
QY 643 CTGACCCAGCTGGGCTGCACCCCTGAATCTCCCATCAGCCCATCGAGACCGTGGCGGTG 702  
DB 141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160  
QY 703 AAGCTGAAGCCCGGCTGAGCGGCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAAG 762  
DB 161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180  
QY 763 ATCAAGCCCTGACCGCCATCTGCGAGGAGATGAGAGGAGGCGCAAGATCACCAAGATC 822  
DB 181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200  
QY 823 GGCCCCGAGACCCCTACACACCCCGTGTTCGCCCATCAAGAGAGGACGACCCCAAG 882  
DB 201 GlyProGluAsnProTyThrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

QY 883 TGGCGCAAGCTGTGGAGCTTCCCGAGCTGAACAAGCGCACCCAGGACTTCTTGGAGGTG 942  
DB 221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240  
QY 943 CAGCTGGGCGCTCCCGCACCCCGCGCTGAAGAAGAAGAAGAGAGCGTGCCTGTGCGAC 1002  
DB 241 GlnLeuGlyIleProHisProAlaIleGlyLeuLysLysLysSerValThrValLeuAsp 260  
QY 1003 GTGGCGCAGCCCTACTTTCAGCGTCCCTCGACGAGGACTTCCGCAAGTACACCGCTTC 1062  
DB 261 ValGlyAspAlaTyrrPheSerValProLeuAspGluGlyPheArgLysTyrrThrAlaPhe 280  
QY 1063 ACCATCCCGAGCATCAACAACAGAGACCCCGCGCATCCGCTACCAAGTACAACGTGTGCC 1122  
DB 281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrrGlnTyrrAsnValLeuPro 300  
QY 1123 CAGGGCTGGAAGGCGACCCCGACGATCTTCAGAGCAGCATGACCAAGATCCTCGAGGCC 1182  
DB 301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGlnPro 320  
QY 1183 TTCCGGCGCGCAACCCCGAGATCGTGATCTACCCAG-----GCCCGCTGTAGCTGGGC 1236  
DB 321 PheArgAlaGlnAsnProGluIleValIleTyrrGlnTyrrMetAspAspLeuTyrrValGly 340  
QY 1237 AGCGACCTGGAGATCGGCGCAGCACCCCGCGCAAGATCGAGGAGCTGCGCAACGACCTGCTG 1296  
DB 341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360  
QY 1297 CGCTGGGGCTTCCACCCCGCAACAGCAGACAGGAGGAGCCCTCTCTCTGTGGATG 1356  
DB 361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380  
QY 1357 GGCTTACGAGCTGCACCCCGCAAGTGCAGCCCTCGAGCCCATCGAGCTGCCCGAAGAGAG 1416  
DB 381 GlyTyrrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400  
QY 1417 AGCTGAGCCGTGAACGATCCAGAACCTGGTGGCAAGCTGAACCTGGCGCAGCAGATC 1476  
DB 401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420  
QY 1477 TACCCGGCATCAGGTGGCGCGAGCTGTGAAGCTGTGCGCGCGCCAGGCGCTGACC 1536  
DB 421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440  
QY 1537 GACATCGTGCCTGACCGGAGGCGCGAGCTGAGCTGCGCGAGAACCGGAGATCTGT 1596  
DB 441 AspIleValProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeu 460  
QY 1597 CGCGAGCCCTGACCGCGCTGTACTACGACCCCAAGCAAGACCTGTGTGGCGCAGATCCAG 1656  
DB 461 LysGluProValHisGlyValTyrrAspProSerLysAspLeuIleAlaGluIleGln 480  
QY 1657 AAGCAGGCGCACCGACCTGAGCTTACAGATCTTACAGAGGCGCTTCAGAACCTGAG 1716  
DB 481 LysGlnGlyAspAspGlnTrpThrTyrrGlnIleTyrrGlnGluProPheLysAsnLeuLys 500  
QY 1717 ACCGGCAAGTACCGCAAGATGCGCACCGCCACACCAACGACGTAAGACGACTGACCGAG 1776  
DB 501 ThrGlyLysTyrrAlaLysArgArgThrHisThrAsnAspValLysGlnLeuThrGlu 520  
QY 1777 GCCGTGCAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCG 1836  
DB 521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540  
QY 1837 CTGCCCATCCAGAGGAGACCTGGGAGACCTGTGGAGCCGACTCTGGCGCCGACCTGG 1896  
DB 541 LeuProIleGlnLysGluThrTrpGluIleTrpThrAspTyrrTrpGlnAlaThrTrp 560  
QY 1897 ATCCCGGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGGTACCACTGAG 1956  
DB 561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrrGlnLeuGlu 580



QY 1957 AAGGAGCCCATCATCGCGCCGAGACCTTACGTGGACGGCGCCGCAACCCGAGACC 2016  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 581 LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600  
 QY 2017 AAGATCGGCAAGCGCGCTACGTGACGACCGCGCGGCGGAGATCGTGACCTGACC 2076  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 601 LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620  
 QY 2077 GAGACACCAACAGACGCGAGCTGCGAGCGCATCCAGCTGGCGCTGCGAGACGGCG 2136  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640  
 QY 2137 AGCGAGTGAACATCGTGACGACGACGACGACGACGACGACGACGACGACGACGAC 2196  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 641 SerGluValAsnIleIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660  
 QY 2197 GACAAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCGAGCTGATCAAGAGGAGA 2256  
 Db |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
 661 AspArgSerGluSerGluLeuValAsnGlnIleIleGlnLeuIleLysLysGluArg 680  
 QY 2257 GTGTACTGAGCTGGTGGCGCCGACAGGCGATCGCGGCGACGACGATCGACAG 2316  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 681 ValTyrLeuSerTrpValProIleHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700  
 QY 2317 CTGCTGAGCAAGGCGCATCCGCAAGCTGCTTCTGAGCGGCATCGAT 2364  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716  
 RESULT 8  
 AAB86169  
 ID AAB86169 standard; protein; 3025 AA.  
 XX  
 AC AAB86169;  
 DT  
 XT 11-SEP-2003 (revised)  
 DT 21-AUG-2001 (first entry)  
 XX  
 XX HIV-1 subtype C protein fragment #1.  
 XX  
 XX Infection; diagnosis; human; humoral immune response; antiviral;  
 KW cellular immune response; vaccine; treatment; gene therapy.  
 KW  
 XX Human immunodeficiency virus 1.  
 XX  
 XX  
 XX  
 PH Key Location/Qualifiers  
 FT Misc-difference 1..3025  
 FT /label= Xaa  
 FT /note= "Xaa represents a stop codon"  
 FT  
 FT  
 PN DB10056747-Al.  
 XX  
 XX  
 PD 31-MAY-2001.  
 XX  
 XX 16-NOV-2000; 2000DE-01056747.  
 XX  
 XX 16-NOV-1999; 99DE-01055089.  
 XX  
 XX (SHAO/) SHAO Y.  
 PA (GENE-) GENEART GMBH GES ANGEWANDTE BIOTECHNOLOG.  
 XX  
 XX Wagner R, Wolf H, Shao Y, Graf M;  
 PI  
 XX WPI; 2001-336417/36.  
 DR N-PSDB; AAH20870.  
 XX  
 XX New nucleic acid sequences from a human immune deficiency virus  
 PT intersubtype, useful for treatment, prevention and diagnosis of  
 PT infection.  
 XX  
 XX Disclosure; Fig 8A-O; 48pp; German.  
 PS  
 XX This invention describes a novel polynucleotide isolated from human  
 CC immunodeficiency viruses type 1 subtype C/B' which can be used for the

CC induction of specific humoral and cellular immune responses. (I) and  
 CC polypeptides (II) encoded by them, are useful in pharmaceuticals,  
 CC vaccines and diagnostic agents, particularly for treatment or prevention  
 CC of human immune deficiency virus-1 (HIV-1) infections, also for rational  
 CC design of test or therapeutic reagents, or gene therapy vectors.  
 CC Polypeptides, especially antibodies, specifically directed against (II)  
 CC are similarly useful as pharmaceutical and diagnostic agents. (I) are  
 CC specific for intersubtype C/B' of HIV-1 so are useful in regions  
 CC (particularly China and South-East Asia) where this subtype is prevalent.  
 CC The products of the invention have antiviral activity. This sequence  
 CC represents a protein encoded by the HIV-1 subtype C genome described in  
 CC the method of the invention. (Updated on 11-SEP-2003 to standardise OS  
 CC field)  
 XX  
 SQ Sequence 3025 AA;  
 Alignment Scores:  
 Pred. No.: 1,95e-197 Length: 3025  
 Score: 3563.00 Matches: 682  
 Percent Similarity: 91.65% Conservative: 42  
 Best Local Similarity: 86.33% Mismatches: 57  
 Query Match: 77.91% Indels: 9  
 DB: 4 Gaps: 5  
 US-09-610-313B-31 (1-2463) x AAB86169 (1-3025)  
 QY 15 GGCAGGAGCCATGAGCCAGCCAGCCAGCCAGCCATCTGATGCGAGCAACTTCAA 74  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 416 GlyArgGlyAsnGluProAsnLysGln---CysHisThrAspAlaGluLysGlnPhe\*\*\* 434  
 QY 75 GGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGCCACATCCCGCCAA 134  
 Db |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
 435 ArgLeu\*\*\*LysAsnCys\*\*\*MetPheGlnLeuTrpGlnGlyArgAlaHisSerGlnLys 454  
 QY 135 CTGCGCGCGCCCGCCCAAGAGGGCTGTGGAAGTGGCGGCAAGAGGGCCACCAGATGAA 194  
 Db |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
 455 LeuGlnGlyPro\*\*\*GluLysGlyLeuLeuGluMetTrpLysArgArgThrProAsnGlu 474  
 QY 195 GGNACTGACCGAGCGCGGCGGCAACTTCTTCGCGAGAGGCTGCGCTCCCGCCAGGCAA 254  
 Db |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
 475 ArgLeuTyr\*\*\*GluThrGlyGln-PhePheArgGluAsnLeuAlaLeuProGlnGlyAr 494  
 QY 255 GCGCCGCGAGTTCCTCCCGAGGAGCAACCGCGCCCAACAGCCCGCCAGCGCGAGTCCA 314  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 494 gAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerProThrArgGlyGluLeuGl 514  
 QY 315 GGTG-----CGCGCGCAACACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCCCT 368  
 Db |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
 514 nValTrpGlyArgAspAsnAsnSerIleSerGluAlaGlyAlaAsnArgGlnGlyThrIl 534  
 QY 369 G-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGAGCATCAAGGTGGG 422  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 534 eSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGl 554  
 QY 423 CGGCCAGATCAAGAGGCGCTCTGTGGACACCGCGCGCGCGCGCACACCGTGTCTGGAGGAGAT 482  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 554 yGlyGlnLeuLysGluAlaLeuAspThrGlyAlaGlyAspThrValLeuGluAspLe 574  
 QY 483 GAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCTTCATCAAGGT 542  
 Db |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
 574 uAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysVa 594  
 QY 543 GCGCCAGTACGACCATCTGATCGAGATCTCGGCAAGAGCCATCGGCGCGCGCGCGCTGCT 602  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 594 lArgGlnTyrGluGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThrValLe 614  
 QY 603 GATCG 662  
 Db |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::  
 614 uValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysTh 634  
 QY 663 CTGGAATCTCCCATCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 722  
 Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 634 rLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAs 654

QY 723 CGGCCCCAAGGTGAAGCAGTGCCCTCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCCAT 782  
 Db 654 pGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIle 674  
 QY 783 CTGGGAGAGATGGAAGAGAGGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACAA 842  
 Db 674 eCyAspGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAs 694  
 QY 843 CACCCCGCTGTTCGCATCAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902  
 Db 694 nThrProIlePheAlaIleLysLysAspSerThrLysIleAspValLeuValAspPh 714  
 QY 903 CGGAGAGCTGAACAGCCACCCAGGACTTCTGGAGGTGAGCTGGGATCCGCCACCC 962  
 Db 714 eArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734  
 QY 963 CGCCGGCTGAAG 1022  
 Db 734 oAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754  
 QY 1023 CGTGCCTGACGAGAGATTCGCGAAGTACACAGCCCTTCCACATCCCGCAGCATCAACA 1082  
 Db 754 rIleProLeuThrGluAspPheArgLysTyrThrAlaPheThrIleProSerArgAsnAs 774  
 QY 1083 CGAGACCCCGGATCCGCTACAGTACAGTCAACGCTGCTGCCCGGAGGTGGAAGGCGAGCC 1142  
 Db 774 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyrLysGlySerLe 794  
 QY 1143 CAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGGCCCTTCGCGCCCGCCACCCCGA 1202  
 Db 794 uAlaIlePheGlnSerSerMetThrLysThrLeuGluProPheArgLysGlnAsnProGl 814  
 QY 1203 GATCGTGATCTACAG-----GCCCCCTGTACGTGGGCGAGCATCTGGAGATCGGCA 1256  
 Db 814 yIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGl 834  
 QY 1257 GCACCGCGCCAGATCGAGGAGCTGCGCAACGACCTGCTGCGCTGGGCTTCACACCC 1316  
 Db 834 nHisArgThrLysIleGluGluLeuArgLysGlnHisLeuLeuArgTyrGlyPheThrThrPr 854  
 QY 1317 CGAAGAGACACAG 1376  
 Db 854 oAspLysLysHis---LysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAs 873  
 QY 1377 CAAGTGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1436  
 Db 873 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerTrpThrValAsnAspIl 893  
 QY 1437 CCAGAGCTGTGGGCAAGCTGAAGTGGGCGAGAGTACCCCGGAGATCAAGGTGCG 1496  
 Db 893 eGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValAr 913  
 QY 1497 CCAGCTGCAAGCTGCTCGCGGCGGCGCAAGCCCTGACGACATCGTGCCTCCCGCAG 1556  
 Db 913 gGlnLeuCysLysLeuLeuArgLysAlaLysAlaLeuThrAspIleValProLeuThrGl 933  
 QY 1557 GGAGCGGAGCTGGAGCTGGCGGAGAACCGGAGATCTCGCGAGAGCCGCGGCGAGCGGT 1616  
 Db 933 uGluAlaGluLeuLeuAlaGluAsnArgLysIleLeuLysGluProValHisGlyVa 953  
 QY 1617 GTACTAGACCCCGAGAGAGCTGTGGCGGAGATCCAGAGAGAGAGAGAGAGAGAGAGAG 1676  
 Db 953 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973  
 QY 1677 GACCTACAGATCTACAGAGAGCCCTTCAAGAACCTGAAGACCGGCGAGTACGCGCAAGAT 1736  
 Db 973 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993  
 QY 1737 GCGCACCGCCACACAG 1796  
 Db 993 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

QY 1797 GGAGAGCATCTGATCTGGGCGAAGACCCCAAGTTCCGCTGCGCCATCCAGAGAGAGAC 1856  
 Db 1013 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033  
 QY 1857 CTGGGAGACCTGGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGT 1916  
 Db 1033 rTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053  
 QY 1917 GAACACCCCGCTGTGTGAAGCTGTGTATCCAGCTGGAGAGAGAGAGAGAGAGAGAG 1976  
 Db 1053 lAsnThrProProLeuValLysLeuTyrTrpGlnLeuGluLysAspProIleValGlyVa 1073  
 QY 1977 CGAGACCTTCTACGTGGAGCGCGCCCAACCGGAGAGAGAGAGAGAGAGAGAGAGAG 2036  
 Db 1073 lGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTy 1093  
 QY 2037 CGTGACCGAGCGCGCGAGAGATCGTGAGCTGACCGAGAGAGAGAGAGAGAGAGAGAG 2096  
 Db 1093 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113  
 QY 2097 CGAGCTGCAGGCCATCCAGCTGGCCCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2156  
 Db 1113 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133  
 QY 2157 CGAGACCGAGTACGCTGGGATCATCCAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 2216  
 Db 1133 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153  
 QY 2217 GGTCAACCGAGATCCAGAGAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2276  
 Db 1153 uValAsnGlnIleIleGluGlnLeuMetLysGluArgValTyrLeuSerTrpValPr 1173  
 QY 2277 CGCCCAAGGCGATCGCGGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2336  
 Db 1173 oAlaHisLysGlyIleGlyLysGlnGluValAsnGluValAspLysLeuValSerSerGlyIleAr 1193  
 QY 2337 CAAGGTGCTGTCTGAGCGGATCGAT 2364  
 Db 1193 gLysValLeuPheLeuAspGlyIleAsp 1202

RESULT 9  
 AA48949  
 ID AA48949 standard; protein; 854 AA.  
 XX  
 AC AA48949;  
 XX  
 DT 29-AUG-2003 (revised)  
 DT 19-APR-2002 (first entry)  
 XX  
 DE HIV-1 subtype C isolate Dul51 pol protein.  
 XX  
 KW HIV-1 subtype C; vaccine; HIV infection; AIDS; pol; antiviral.  
 XX  
 OS Human immunodeficiency virus; type I.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 58 /note= "encoded by TGA"  
 FT Misc-difference 72 /note= "encoded by TAA"  
 FT Misc-difference 114 /note= "encoded by TAA"  
 FT Misc-difference 117 /note= "encoded by TAA"  
 FT  
 XX MO200204494-A2.  
 XX  
 PD 17-JAN-2002.  
 XX  
 PF 09-JUL-2001; 2001WO-IB001208.  
 XX  
 PR 07-JUL-2000; 2000US-0216995P.  
 PR 10-JUL-2000; 2000ZA-00003437.

```
PR 15-SEP-2000; 2000ZA-00004924.
XX (MEDI-) MEDICAL RES COUNCIL.
PA (UYCA-) UNIV CAPE TOWN.
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Williamson C, Swannstrom RI, Morris L, Karim SA, Johnston RE;
XX WPI; 2002-171700/22.
DR N-PSDB; AAL41592.
XX
XX Selecting HIV-1 subtype C isolates, which are useful in developing
PT vaccines against HIV infection, comprises isolating viruses with high
PT sequence identity to a consensus sequence whose phenotype is associated
PT with the HIV subtype.
XX
XX Claim 19; Page 67; 69pp; English.
XX
XX The present invention relates to a process for the selection of human
CC immunodeficiency virus (HIV) subtype isolates for use in the development
CC of a prophylactic and/or therapeutic pharmaceutical composition. The
CC process involves selecting isolated virus or viruses with a high sequence
CC identity to a consensus sequence and a phenotype which is associated with
CC transmission for the particular HIV subtype. The composition can be used
CC in the production of vaccines against HIV. The present sequence is the
CC HIV-1 subtype C isolate DUL51 pol protein. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
XX SQ Sequence 854 AA;
XX
Alignment Scores:
Pred. No.: 3 94e-195 Length: 854
Score: 3521.00 Matches: 666
Percent Similarity: 94.31% Conservat: 30
Best Local Similarity: 90.24% Mismatches: 29
Query Match: 77.00% Indels: 13
DB: 5 Gaps: 3
XX
US-09-610-313B-31 (1-2463) x AAM48949 (1-854)
QY 171 CGGCAAGGAGGCGCCAGATGAGGACTGCACCGAGCGCCAGGCCAAGTCTT----- 224
DB 118 ArgGlnGlyPheProSerHisAppValValLysArgArgProValProSerLeuHisAla 137
QY 225 -----CCGCGAGGACCTGGCC--TTCCCGGAGGCGAAGCCCG 260
DB 138 CysArgSerThrLeuGluAspProArgValPro-SerSerPheProGlnGlyProAla 157
QY 261 CGAGTTCCCGAGGAGCAGAACCGCGCAACAGCCCGCCAGCGCGCGAGCTGCGAGTGC 320
DB 157 GlnPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeuGlnVal 177
QY 321 CGGCGCAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
DB 177 gArgAspAsnProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAsnPhePro 197
QY 381 GATCACCTGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440
DB 197 nileThrLeuTrpGlnArgProLeuValSerIleLysIleGlyGlnThrArgGluAl 217
QY 441 CTGCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
DB 217 aLeuLeuAspThrGlyAlaAspThrValLeuGluAspIleAsnLeuProGlyLysTr 237
QY 501 GAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
DB 237 pLysProLysMetIleGlyGlyGlyGlyPheIleLysValArgGlnTrpAspGln 257
QY 561 CTGATCGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620
DB 257 eLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValProThrPro 277
QY 621 GAACATCATCGCGCGCAACATGTCACCGCGCTGGCGTGCACCTGGAACCTTCC 680
277 lAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSe 297
681 CCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGGAGCGGCCCAAGGTGAAGCA 740
297 rProfileGluThrValProValLysLeuLysProGlyMetAspGlyProLysValysG 317
741 GTGGCCCTGACCGAGGAGAGATCAAGCCCTGACCGCATCTCGAGAGAGATGAGAGAA 800
317 nTrpProLeuThrGluValLysIleLysAlaLeuThrAlaIleCysGluGluMetGlu 337
801 GAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTCAACACCCCGTGTTCGCCAT 860
337 sGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIlePheAla 357
861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGAGGTGAACAGCG 920
357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 377
921 CACCGAGGACTTCTGGGAGGTGAGCTGGCGCATCCCGCCACCCCGCGCGCTGAAGAGAA 980
377 gThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLys 397
981 GAAGACGCTGACCGTCTCGACGTGGCGCGCGCGCGCTACTTTCAGCGTGCCTGACGAG 1040
397 sLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGlu 417
1041 CTTCCGCAAGTACACCGCTTACCATCCCGCGCATCAACAGAGACCCCGCGCATCCG 1100
417 yPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIle 437
1101 CTACCAAGTACACGCTGCTGCGCCCGAGGCGTGAAGGCGAGCGCCAGCATCTTCCAGAG 1160
437 gTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAla 457
1161 CATGACCAAGATCCTCGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1218
457 rMetThrLysIleLeuGluProPheArgAlaLysAsnProGluIleValIleTyrGln 477
1219 -----GCCCGCTGTAGTGGCGAGCGACCTGGAGATCGCGCGCGCGCGCGAGATCGA 1274
477 rMetAlaAlaLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 497
1275 GAGCTGCGCGAAGCACCTGCTGCGCTGGGGCTTTCACACCGCGCGCGCGCGCGCGAG 1334
497 uGluLeuArgGluHisLeuLysTrpGlyPheThrThrProAspLysLysHisGlnL 517
1335 GAGCGCGCGCTTCTCTGTGATGGGTACGAGCTGCGACCGCGCGCGCGCGCGCGCGCG 1394
517 sGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGln 537
1395 CATCGAGCTGCCGAGAGAGAGCTGGACCGGTGAACGACATCCAGAGCTGTGTGGCGAA 1454
537 oIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 557
1455 GCTGAATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1514
557 sLeuAsnTrpThrSerGlnIleTyrProGlyLysValArgGlnLeuCysLysLeuLe 577
1515 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1574
577 uArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluAlaGluLeuGluLe 597
1575 GCGCGAGAACCGGAGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1634
597 uAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAspProSer 617
1635 GGAACCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1694
617 sAspLeuIleAlaGluIleGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 637
1695 GGAGCGCTTCAAGAACCTGGAAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1754
```

Db 637 nGluProPheLysAsnLeuLysThrGlyLysTyAlaLysArgThrThrHisThrAs 657  
 Qy 1755 CGAGCTGAAGCAGCTGACCGAGCGCGTCACAGAGATCCCATGGAGAGCATCGTGATCTG 1814  
 Db 657 nAspValLysGlnLeuThrGluAlaValGlnLysLeuSerLeuGluSerLeuValThrTr 677  
 Qy 1815 GGGCAAGACCCCAAGTTCGCTGCCCTGCCATCCAGAGGAGACCTGGAGACCTGGTGAC 1874  
 Db 677 pGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluIleThrTrpTh 697  
 Qy 1875 CGACTACTGGCAGCGCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGT 1934  
 Db 697 rAspTyTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVa 717  
 Qy 1935 GAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCCGAGACCTTTCTACGTGGA 1994  
 Db 717 lLysLeuTrpTyTrpGlnLeuGluLysGluProIleAlaGlyAlaGluThrPheTyValAs 737  
 Qy 1995 CGGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGTACGTGACCGACCGGGCGG 2054  
 Db 737 pGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyValThrAspArgGlyAr 757  
 Qy 2055 GCAGAAGATCTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGGCCATCCA 2114  
 Db 757 gGlnLysIleValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGI 777  
 Qy 2115 GCTGCGCCTCGAGACAGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174  
 Db 777 nLeuAlaLeuGlnAspSerGluSerGluValAsnIleValThrAspSerGlnTyAlaLe 797  
 Qy 2175 GGGCATATTCAGGCCCGCAGCCGACAGAGCGAGCGAGCGAGTGTGTGAACCGATCATCGA 2234  
 Db 797 uGlyIleIleGlnAlaGlnProAspArgSerGluSerGluLeuValAsnGlnIleIleGI 817  
 Qy 2235 GCAGCTCATCAAGAGGAGAGGTGTACCTGAGCTGGTGCCCGCCACAGGGGCATCGG 2294  
 Db 817 uGlnLeuLysLysGluArgAlaTyLeuSerTrpValProAlaHisLysGlyIleGI 837  
 Qy 2295 CGGCAAGCAGCAGATGACAGCTGGTGAGCAAGGGCATCGCGAAGGTGCTG 2346  
 Db 837 yGlyAspGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 854

RESULT 10  
 AAB69282  
 ID AAB69282 standard; protein; 1000 AA.  
 XX AC AAB69282;  
 XX DT 12-SEP-2003 (revised)  
 XX DT 20-APR-2001 (first entry)  
 XX DE HIV-1 non-subtype B clone 92RW009-6 pol protein.  
 XX KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 XX KW vif; vpr; tat; rev; nef; vaccine.  
 XX OS Human immunodeficiency virus 1.  
 XX PN WO20026416-A1.  
 XX PD 11-MAY-2000.  
 XX XX 25-OCT-1999; 99WO-US024837.  
 XX XX 02-NOV-1998; 98US-00184418.  
 XX XX (UABR-) UAB RES FOUND.  
 XX XX Hahn BH, Shaw GM, Gao F;  
 XX XX WFI; 2000-365651/31.  
 XX PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.

PS Claim 41; Fig 15; 131pp; English.

XX The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX SQ Sequence 1000 AA;

Alignment Scores:  
 Pred. No.: 6,5e-195 Length: 1000  
 Score: 3517.50 Matches: 660  
 Percent Similarity: 95.96% Conservative: 29  
 Best Local Similarity: 91.92% Mismatches: 26  
 Query Match: 76.92% Indels: 3  
 Gaps: 2

US-09-610-313B-31 (1-2463) x AAB69282 (1-1000)

Qy 220 TTCTTCGCGAGGACCTGGCTTCCCCAGGCAAGCGCCGCGATTCCCCAGCGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20  
 Qy 280 AACCGCCCAACACCCCGCAGCGCGAGCTGCAGGTGCGCGCGC---GACACCCCGCGC 336  
 Db 21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyGlyArgAspSerLeuSer 40  
 Qy 337 AGCGAGCGCGCGCGAGCGCGAGCGCACCTGAACCTTCCCCAGATCACCCTGTGGCAG 396  
 Db 41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnIleThrLeuTrpGln 60  
 Qy 397 CGCCCCCTGTGTGACATCAAGGTGGCGCGCGAGATCAAGAGAGCGCCTGTGGACACCGCGC 456  
 Db 61 ArgProLeuValThrValLysIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80  
 Qy 457 GCGACGACACCGCTGTGGAGGAGATGAGCTGCCCGCAAGTGAGAGCCCAAGATGATC 516  
 Db 81 AlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMethIle 100  
 Qy 517 GCGGCGATCGCGCTTCATCAAGTGGCGCGAGTACGACAGATCCTGTGATCGAGATCTGC 576  
 Db 101 GlyGlyIleGlyGlyPheIleLysValLysGlnTrpAspGlnIleLeuIleGluIleCys 120  
 Qy 577 GCGAAGAAGCGCATCGCGACCGCTGTGATCGCGCCCGCACCCTCGTGAACATCATCGCGCGC 636  
 Db 121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleGlyArg 140  
 Qy 637 AACATGTGACCGAGCTGGGCTGCACCTGAACCTTCCCATCAGCCCATCGAGACCGGTG 696  
 Db 141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160  
 Qy 697 CCGTGAAGTGAAGCCCGCGCATCGCGCGCGCGAGGTGAAGAGTGAAGCGCTGACCGCGAG 756  
 Db 161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 180  
 Qy 757 GAGAAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGATGGAGAAGAGGGGCAAGATCAC 816  
 Db 181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluLysLysIleSer 200  
 Qy 817 AAGATCGCGCGCGAGAACCCCTACAAACCCCGCTGTTCGCCATCAGAGAAGAGCAGCAGC 876  
 Db 201 LysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLysLysAspSer 220  
 Qy 877 ACCAAGTGGCGAGCTGGTGGACTTCGCGAGCTGAACAGCGCACCGAGCTCTCGG 936  
 Db 221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240

QY 937 GAGTGCAGCTGGCGATCCCCACCCCGCGGCTGAAGAAGAAGACGCGTGACCGTG 996  
Db 241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysValThrVal 260  
QY 997 CTGGACCTGGCGCAGCGCTACTTTCAGCGTCCCTCGACGAGACTTCGCAAGTACACC 1056  
Db 261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280  
QY 1057 GCCTTACCATCCCGACATCAACACGAGACCCCGGATCCGCTACGATGACACGCTG 1116  
Db 281 AlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 300  
QY 1117 CTGCCCCAGCGCTGGAGGCGACCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTG 1176  
Db 301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320  
QY 1177 GAGCCCTTCGCGCGCCCAACCCCGAGATCGTATCTACACAG-----GCCCCCTGTAC 1230  
Db 321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspAspLeuTyr 340  
QY 1231 GTGGGACGACCTGGAGATCGGCGCAGCACCGCGCCMAAGATCGAGAGCTGCGCAAGCAC 1290  
Db 341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHis 360  
QY 1291 CTGCTGCGCTGGGCTTCACACCCCGCACAGAGCACCAGAGGAGCCCTTCCTG 1350  
Db 361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeu 380  
QY 1351 TGGATGGGCTACGAGCTGCAACCCCGACCAAGTGACCGTCCAGCCCATCGAGCTGCCCGAG 1410  
Db 381 TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400  
QY 1411 AAGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGC 1470  
Db 401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420  
QY 1471 CAGATCTACCCCGCATAGTGCAGCTGCAGCTGCAAGCTGCTGCGCGCGCCAGGCC 1530  
Db 421 GlnIleTyrProGlyValLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAla 440  
QY 1531 CTGACCCACATCGTGCCTCCGACGAGGAGCGGAGCTGGAGCTGGCCGAGAACCCGCGAG 1590  
Db 441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGlu 460  
QY 1591 ATCTGCGCGAGCCGTGCACGGCTGTACTAGACCCCGACGAGACCTGTGTGGCGGAG 1650  
Db 461 IleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGlu 480  
QY 1651 ATCCAGAGCAGGCGCCACGACCTGAGCTGACCATCTACAGAGGAGCCCTTCAAGAAC 1710  
Db 481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500  
QY 1711 CTGAAGACCGGCAAGTACGCGCATCGCACCGCCCGCACCAACGACGCTGAAGCAGCTG 1770  
Db 501 LeuLysThrGlyLysTyrAlaLysArgArgThrAlaHisThrAsnAspValLysGlnLeu 520  
QY 1771 ACCGAGCCGTGCAGAAAGATCGCCATGGAGAGCATCGTATCTGGGGCAAGACCCCAAG 1830  
Db 521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrpGlyLysThrProLys 540  
QY 1831 TTCGCTGCTCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCC 1890  
Db 541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAla 560  
QY 1891 ACCTGATCCCGAGTGGGAGTTCTGTAACACCCCGCTGTGTGAAGCTGTGTACACAG 1950  
Db 561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580  
QY 1951 CTGGAAGAGGACCATCATCTGCGCGCGCAGACCTTCTACGTGAGCGCGCGCCCAACCGC 2010  
Db 581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

QY 2011 GAGACCAAGATCGGCAAGCGCCGCTACGTACACGACCGGCGCGCAGAGATCGTGAGC 2070  
Db 601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620  
QY 2071 CTGACCGAGACCAACCAAGAACCGAGCTGCGAGCCATCCAGCTGGCCCTGCAGGAC 2130  
Db 621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640  
QY 2131 AGCGGCGAGCGGTGAACATCGTGACCGACGACGAGCTAGCCCTGGGCATCATCCAGGCC 2190  
Db 641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 660  
QY 2191 CAGCCCGCAAGACGAGCGAGCTGCTGGAACACGAGATCATCCAGAGCTGATCAAGAAG 2250  
Db 661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleIleGluGlnLeuIleLysLys 680  
QY 2251 GAGAAAGTGTACGTAGCTGGTGGTGGCCGCCCAAGGGCATCGCGGCGCAACGAGCAGATC 2310  
Db 681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 700  
QY 2311 GACAAGCTGGTGAGCGGCGCATCCGCAAGGTGCTGTCTCGACGCGCATCGAT 2364  
Db 701 AspLysLeuValSerSerGlyIleArgValLeuPheLeuAspGlyIleAsp 718  
RESULT 11  
AAW72993  
ID AAW72993 standard; protein; 1002 AA.  
XX AC AAW72993;  
XX 25-MAR-2003 (revised)  
DT 15-FEB-1999 (first entry)  
XX DE HIV isolate LAV.MAL pol protein.  
XX LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;  
KW pol.  
XX Human T-lymphotropic virus.  
XX OS  
XX PN US5824482-A.  
XX PD 20-OCT-1998.  
XX PF 06-JUN-1995; 95US-00471474.  
XX PR 23-JUN-1986; 86FR-00040138.  
PR 13-APR-1987; 87US-00038330.  
PR 19-FEB-1991; 91US-00656797.  
PR 10-DEC-1992; 92US-00988530.  
PR 18-NOV-1993; 93US-00154397.  
XX (INSP ) INST PASTEUR.  
XX PA  
XX Wain-Hobson S, Sonigo P, Alizon M, Montagnier L;  
XX WPI; 1998-582548/49.  
XX N-PSDB; AAV63467.  
XX Human immunodeficiency virus isolate LAV(MAL) - and method for detecting  
XX anti-HIV antibodies.  
XX PS Disclosure; Fig 7B-E; 47pp; English.  
XX CC This is the amino acid sequence of the pol protein of lymphadenopathy  
CC associated virus LAV.MAL (NCIM 1-641), a new virus isolate from Zaïre  
CC that is responsible for diseases clinically related to AIDS. The sequence  
CC was deduced from an open reading frame (ORF) of the LAV.MAL genome (see  
CC AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of  
CC the envelope glycoprotein can be used as antigens in a claimed method for  
CC the in vitro detection of an antibody directed against LAV. The method is  
CC useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in  
CC patients, asymptomatic carriers and in blood-related products. LAV viral

CC antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct PR field.)

XX  
SQ Sequence 1002 AA;

Alignment Scores:  
Pred. No.: 7,14e-194 Length: 1002  
Score: 3499.50 Matches: 653  
Percent Similarity: 96.11% Conservative: 39  
Best Local Similarity: 90.69% Mismatches: 23  
Query Match: 76.53% Indels: 5  
DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x AAW72993 (1-1002)

Qy 220 TTCTTCGGCAGGACCTGGCCCTTCCCGCAGGCGAAGGCCCGCGAGTTCCCGCAGCGAGCAG 279  
Db 1 PhePheArgGluAenLeuAlaPheProGlnGlyAlaArgGluPheProSerSerGluGln 20  
Qy 280 AACCGCGCAACAGCCCGCAGCGGAGCTGAGGTG---CGCGGCGCAACCCCGCGC 336  
Db 21 ThrArgAlaAenSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40  
Qy 337 AGCGAGCGCGCGCGAGCGCGAGGC-----ACCCTGAACCTTCCCGCAGATCACCTG 390  
Db 41 SerGluThrGlyAlaGluArgGlnGlyLeValSerPheSerPheProGlnThrLeu 60  
Qy 391 TGGCAGCGCCCGCTGGTGAGCATCAAGGTGGCGCGCAGATCAAGAGGCGCCCTCTCTGAC 450  
Db 61 TrpGlnArgProValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80  
Qy 451 ACCGCGCGCAGCACCGTCTCGAGAGATGAGCTGCCCGCAAGTGAAGCCCAAG 510  
Db 81 ThrGlyAlaAspAspThrValLeuGluGluLeuAsnLeuProGlyLysTrpLysProLys 100  
Qy 511 ATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCAGTACGACGAGATCCTGATCGAG 570  
Db 101 MetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnLeuLeuIleGlu 120  
Qy 571 ATCTCGCGCAAGAGGCCATCGGACCGCTGATCGCGCCCGCCCGCGTGAACATCATC 630  
Db 121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleLeu 140  
Qy 631 GGCGCAACATGTGACCCAGCTGGTGCACCTGAACTTCCCATCAGCCCGCATCGAG 690  
Db 141 GlyArgAenMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
Qy 691 ACCGTGCGGTGAGCTGAAGCCCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCTG 750  
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180  
Qy 751 ACCGAGGAGAAGATCAAGGCCCTCACCGCATCTGCGAGGAGATGGAGAGGAGGCGCAAG 810  
Db 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200  
Qy 811 ATCACCAGATCGCGCGCGAGAACCCCTACAAACACCCCGTGTTCGCGCATCAAGAAGAAG 870  
Db 201 IleLeuLysIleGlyProGluAenProTrpAsnThrProValPheAlaIleLysLysLys 220  
Qy 871 GACAGCACCAAGTGGCGCGAGCTGGTGGACTTCGCGAGCTGAACAAGCGCGCACCGAGAC 930  
Db 221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240  
Qy 931 TTCTGGGAGGTGAGCTGGGCGATCCCCACCGCGCGCGCTGAAGAAGAAGAAGAGCGTG 990  
Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
Qy 991 ACCGTGCTGAGCTGGCGAGCGCTACTTTCAGCTGCGCCCTCGGACGAGGACTTCCGCAAG 1050  
Db 261 ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluAspPheArgLys 280  
Qy 1051 TACACCGCTTACCATCCCGCAGCATCAACAGAGACCCCGCGCATCCGCTACCGATAC 1110

Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300  
Qy 1111 AACGTGCTGCCCCAGGGCTGGAAGCGCAGCGCCCGCATCTTCCAGAGCAGCATGACCAAG 1170  
Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
Qy 1171 ATCTGAGAGCCTTCCGGCGCGCAACCCCGAGATCGTGATCTACCAG-----GCCGCC 1224  
Db 321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
Qy 1225 CTGTACGTGGCGCAGCGACCTGAGATCGCGCAGCAGCGCGCAAGATCGAGAGCTGCGC 1284  
Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360  
Qy 1285 AAGCACCTGTGCTGGGCTTCCACCAACCCCGCATCCAGAAAGACCAAGAGGAGCGCCCC 1344  
Db 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
Qy 1345 TTCTGTGATGGCTAGGAGCTGCACCCCGACAAGTGGACCGTGCACCCCATCGAGCTG 1404  
Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
Qy 1405 CCGGAGAGGAGAGCTGCAGCGCTGCAACACATCCAGAAAGCTGGTGGGCAAGCTGAACCTG 1464  
Db 401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
Qy 1465 GCCAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTCTCGCGCGCGCC 1524  
Db 421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440  
Qy 1525 AAGCCCTGACCGACATCGTGCCTCGACCGAGGAGCGCGAGCTGGAGCTGGCGAGAAC 1584  
Db 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460  
Qy 1585 CGCGAGATCTGCGCGAGCGCGCTGACGGGTGCTACTACGACCCCGCAGCAAGCAAGCTGCTG 1644  
Db 461 ArgGluLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
Qy 1645 GCCGAGATCCAGAGAGCGCGCACGACGATGGAGCTACACGATCTACGAGAGCGCTTC 1704  
Db 481 AlaGluIleGlnLysGlnGlyGlnTrpThrTrpThrTyrGlnIleTyrGlnGluTyr 500  
Qy 1705 AAGAACTGAAGACCGCAAGTAGCCCAAGATGCGCACCGCGCCACACCAACGAGCTGAAG 1764  
Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
Qy 1765 CAGTGCACCGCGCGTGCAGAGATCGCATCGAGAGCATCGTGTACTGCGGCAAGAC 1824  
Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
Qy 1825 CCCAAGTTCCGCTGCGCATCCAGAGGAGACCTGGGAGAGCTGGTGGACCGACTACTGG 1884  
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560  
Qy 1885 CAGCCACCTGATGCCGAGTGGAGTTCTGTGAACACCCCGCTGGTGAAGCTGG 1944  
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
Qy 1945 TACCAGCTGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC 2004  
Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
Qy 2005 AACCGCAGAGCAAGATCGCGAGCGCGCTTACGTGACCGACCGCGCGCGCGCGAGATC 2064  
Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
Qy 2065 GTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGCGCATCCAGCTGGCCCTG 2124  
Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
Qy 2125 CAGGACGCGCGCAGGAGTGAACATCGTACCGACGAGCCAGTACGCTGGGCTGATCATC 2184  
Db 641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660



Qy	2185	CAGGCCAGCCCGACCAAGACGAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATC	2244
Db	661	GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleLeuGluLeuLeu	680
Qy	2245	AAGAAGAGAGAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATCGCGGCAACGAG	2304
Db	681	GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyAsnGlu	700
Qy	2305	CAGATCGACAAGCTGTGTGACAAAGGGCATCCGCAAGGTGCTGTTCTTGACGGCATCGAT	2364
Db	701	GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp	720
RESULT 12			
AAO30963	ID AAO30963 standard; protein; 1003 AA.		
XX	AC	AAO30963;	
XX	DT	06-OCT-2003 (first entry)	
XX	XX	HIV pol protein.	
XX	DE		
XX	KW	Human immunodeficiency virus; HIV; pol; IV9; immunotherapy; infection;	
XX	KW	epitope; vaccine; enzyme.	
XX	OS	Human immunodeficiency virus.	
XX	XX		
XX	PN	WO2003051285-A2.	
XX	XX		
XX	PD	26-JUN-2003.	
XX	XX		
XX	PF	29-OCT-2002; 2002WO-US034620.	
XX	XX		
XX	PR	29-OCT-2001; 2001US-0345116P.	
XX	XX		
XX	PA	(GENZ ) GENZYME CORP.	
XX	PA	(GENO ) GEN HOSPITAL CORP DBA MASSACHUSETTS GEN.	
XX	XX		
XX	PI	Nicolette CA, Walker BD;	
XX	XX	WPI; 2003-541603/51.	
XX	DR		
XX	XX		
XX	PT	Novel synthetic therapeutic peptide compounds useful for inducing an	
XX	PT	immune response in a mammal, specifically cross-react with HIV IV9	
XX	PT	epitopes.	
XX	PS	Claim 8; Page 65-67; 67pp; English.	
XX	XX		
XX	CC	The invention relates to synthetic therapeutic peptide compounds that	
XX	CC	specifically cross-react with HIV (human immunodeficiency virus) pol	
XX	CC	protein IV9 epitopes. The invention is useful for treating HIV infection.	
XX	CC	The invention is also used in immunotherapy and as vaccines. The present	
XX	CC	sequence is HIV pol protein	
XX	XX		
SQ	Sequence 1003 AA;		
Alignment Scores:			
Pred. No.:	9.96e-194	Length:	1003
Score:	3497.00	Matches:	655
Percent Similarity:	95.56%	Conservative:	34
Best Local Similarity:	90.85%	Mismatches:	26
Query Match:	76.47%	Indels:	6
DB:	6	Gaps:	3
US-09-610-313B-31 (1-2463) x AAO30963 (1-1003)			
Qy	220	TTCTTCGGCAGGACCTGGGCTTCCCGCAGGCAAGCGCCGCGAGTTCCCGCAGCGAGCAG	279
Db	1	PhePheArgGlnAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
Qy	280	AACCGCGCCAAACGCCCCACCGAGCTGCAGGTG-----CGCGCGCAACACCCC	333

Db	21	ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer	40
QY	334	CGCAGCGAGCGCGCCGAGCGCCAGGCGCACCCCTG-----AATCTCCCGAGATCAC	387
Db	41	ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr	60
QY	388	CTGTGTCAGCGCCCTCGTGTGAGCATCAAGGTGGCGGCAGATCAAGGAGGCCCTGCTG	447
Db	61	LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu	80
QY	448	GACACCGCGCCGACACACCGTGTGTGAGGAGATGAGCTGCCCGCAGGTGGAAGCC	507
Db	81	AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro	100
QY	508	AGATGATCGCGGCATCGCGCTTCATCAGGTGGCGCAGTACGACACAGATCCTGATC	567
Db	101	LysMetIleGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleLeuLeu	120
QY	568	GAGATCTGCGGCAAGAGCCCATCGGCACCGTCTGATCGGCCCCACCCCGTGAACATC	627
Db	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
QY	628	ATCGCGCCGAAACATGCTGACCCAGCTGGGTGACCCCTGACCTGAACTCCCATCAGC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
QY	688	GAGACCGTCCCGTGAAGCTGAAGCCCGCATCGAGCGCCCAAGGTGAAGCAGTGCC	747
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
QY	748	CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAAGAGGGC	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
QY	808	AGATCACCAGATCGCGCCCGAGAACCCCTACAACAACCCCGTGTGTCGCATCAAGA	867
Db	201	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	220
QY	868	AGGACAGCACCAAGTGGCGCAGCTGTGGAGCTTCCGGAGCTGAACAGCCACCCAG	927
Db	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
QY	928	GACTTCTGGAGGTGCAGCTGGGCATCCCCACCCCGCGCTGAAGAGAAGAGAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
QY	988	GTGACCGTGTGACGTGGCGCAGCTACTTACGTGCTGCGGCCCTGACGAGGATTC	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg	280
QY	1048	AGATACACCGCTTCACCATCCCGCAGCATCAACAACGAGACCCCGCATCCGCTACC	1107
Db	281	LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln	300
QY	1108	TACAACGTGTGCTGCCCGCAGGCTGGAAGGGCAGCCCGCATCTTCCAGCAGCATGACC	1167
Db	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320
QY	1168	AGATCTCTGGAGCCCTTCGCGCCCGCACCCGAGATCGTGATCTACACG-----GCC	1221
Db	321	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp	340
QY	1222	CCCTGTGAGTGGCGCAGCACCTGGAGATCGCGCAGCACCGCCCGCAGATCGAGGAGCTG	1281
Db	341	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuLeu	360
QY	1282	CGCAAGCACCTCTGCTGGCGCTTCCACACCCCGCACAAAGACACCAAGAGAGGCC	1341
Db	361	ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro	380
QY	1342	CCCTTCTGTGATGGCTACGAGCTGCACCCCGACAGTGGAGCCCTGCGCCCATCGAG	1401
Db	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal	400

Qy	1402	CTGCCGAGAAAGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGTGGGCAAGCTGAAC	1461
Db	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
Qy	1462	TGGSCCAGCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAGCTGCTCGAGCTCTCGCGGC	1521
Db	421	TrpAlaSerGlnIleIleTyrProGlyIleLysValArgGlnLeuIleTyrLysLeuLeuArgGly	440
Qy	1522	GCCAAAGCCCTGACCGACATCTGTCGCCCTGACCCGAGGAGCGCAGCTGTGAGCTGGCCGAG	1581
Db	441	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu	460
Qy	1582	AACCGCAGATCTCGCCGAGCCCGTGCAGCGCGTGTACTACGACCCCAAGACGACCTG	1641
Db	461	AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu	480
Qy	1642	GTGSCCAGATCCAGAGCAGGGCCACGACCGAGTGACCTACCAGATCTACCAAGGAGCCC	1701
Db	481	IleAlaGluIleGlnLysGlnGlyGlnIleTrpThrTyrGlnIleIleTyrGlnGluPro	500
Qy	1702	TTCAAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGCCCAACCAACACGACGTG	1761
Db	501	PhelYasnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal	520
Qy	1762	AAGCAGCTGACGAGGCGCTGCGAAGATCGCCATGGAGAGCATCTGTGATCTCGGGCAAG	1821
Db	521	LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys	540
Qy	1822	ACCCCAAGTTCGCGCTGCCCATCCAGAAGAGACCTGGGAGACCTGGTCGACCGACTAC	1881
Db	541	ThrProLysPhelLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr	560
Qy	1882	TGGCAGGCCACCTGGATCCCGAGTGGGAGTCTGTGAACACCCCGCTGTGTGAAGCTG	1941
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu	580
Qy	1942	TGTACACGCTGGAGAGGACCCATCATCGCGCCGAGACCTTCTACGTGGACGGCGCC	2001
Db	581	TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla	600
Qy	2002	GCCNACCGCAGACCAAGATCGCAGGCGGCTACGTGACCGACCGCGGCGCGCGAGAG	2061
Db	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys	620
Qy	2062	ATCGTGAACCTGACCCAGACACCAACCAAGACCGAGCTGCAGGCCATCTACGTGGCC	2121
Db	621	ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla	640
Qy	2122	CTCAGGACCGCGCAGCGAGGTGAACATCGTGACCGACCGACCGACCTATCGGGCATC	2181
Db	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle	660
Qy	2182	ATCAGGCCCGCAGCCAGACGAGGAGCGAGCTGTGTGAACCATCATCTCGAGCAGCTG	2241
Db	661	IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu	680
Qy	2242	ATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCCCAAGGCGCATCTCGCGCGCAAC	2301
Db	681	IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn	700
Qy	2302	GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGTGTTCTCTGGACGGCATC	2361
Db	701	GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle	720
Qy	2362	GAT	2364
Db	721	Asp	721

RESULT 13  
ADN36413  
ID ADN3  
XX

AC	ADN36413;
XX	
DT	15-JUL-2004 (first entry)
XX	
DE	HIV protein for anti-HIV vaccine.
XX	
KW	anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX	
OS	Human immunodeficiency virus 1.
XX	
PN	WO2004035006-A2.
XX	
PD	29-APR-2004.
XX	
PF	17-OCT-2003; 2003WO-US033112.
XX	
PR	18-OCT-2002; 2002US-0419465P.
XX	
PA	(AARO-) AARON DIAMOND AIDS RES CENT.
XX	
PI	Huang Y, Ho DD, Chen Z;
XX	
DR	WPI; 2004-348328/32.
XX	
PT	N-PSDB; ADN36412.
XX	
FT	Nucleic acid vector comprising at least one HIV sequence operably linked
PT	to a promoter and encoding a protein that does not assemble into viral
PT	particles, useful in immunizing a subject against HIV infection.
XX	
PS	Disclosure; SEQ ID NO 27; 166pp; English.
XX	
CC	The invention relates to a nucleic acid vector comprising at least one
CC	HIV sequence operably linked to a promoter and encoding a protein that
CC	does not assemble into viral particles. The nucleic acid vector is useful
CC	in immunizing a subject against HIV infection. This sequence corresponds
CC	to a nucleic acid used in the invention.
XX	
SQ	Sequence 1003 AA;
Alignment Scores:	
Pred. No.:	9.96e-194 Length: 1003
Score:	3497.00 Matches: 655
Percent Similarity:	95.56% Conservative: 34
Best Local Similarity:	90.85% Mismatches: 26
Query Match:	76.47% Indels: 6
DB:	8 Gaps: 3

US-09-610-313B-31 (1-2463) x ADN36413 (1-1003)

Qy	220	TTCTTCGCGAGGACCTGGCCCTTCCCCCAGGCAGGCCCGGAGTTCCTCCACGCGAGCAG 279
Dd	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
Qy	280	AACCGCGCCCAACAGCCCCACCAGCGCGAGTGTCAGGTG-----CGCGGCGACAACCCCC 333
Dd	21	ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
Qy	334	CGACGCGAGCGCGCGCGCGCGCGCGCGCACCCCTG-----AACTTCCCCCAGATCAC 397
Dd	41	ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
Qy	388	CTGTGGCAGCGCCCCCTGGTCAGCATCAAGGTGGCGCGCCAGATCAAGAGAGCCCTGCTG 447
Dd	61	LeutrpGlnargProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80
Qy	448	GACACCGCGCGCGACGACACCGTGTGTGAGGAGATGAGCCCTGCCCCGCGGAAGTGGAGGCC 507
Dd	81	AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100
Qy	508	AAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGCGCCAGTAGCACAGATCCTGTATC 567
Dd	101	LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIle 120

QY	568	GAGATCTGCGCGAAGAGGCCATCTGGGCACCGTGTCTGATCGCGCCGCCACCCCGCTGGAACATC	627
Db	121	GlurileCysglyHisIysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
QY	628	ATCGGCGCAACATGCTGACCCAGCTGGGTGCACCTGAACCTCCCATCAGCAGCCCATC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
QY	688	GAGACCGTCCCGTGAAGCTGAAGCCCGGCATCGAGCGGCCCAAGCTGAAGCATGTGCCCC	747
Db	161	GluThrValProValIysLeuLysProGlyMetAspGlyProIysValIysGlnTrpPro	180
QY	748	CTGACCGAGGAGAAATCAAGGCCCTGACCGCCCATCTGCAGGAGAGATGGAAGAAGAGGGCC	807
Db	181	LeuThrGluGluIysIleLysAlaLeuValGluIleCysThrGluMetGluIysGluGly	200
QY	808	AAGATCACCAAGATCGGCCCCGGAAGACCCCTACAACACCCCGCTGTTCGCCATCAAGAAG	867
Db	201	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	220
QY	868	AAGCAGCAGCACCAAGTGGCGCAAGCTGGTGGCACTTCGCGAGCTGAAACAAGCCACCCAG	927
Db	221	LysAspSerThrLysTrpLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
QY	928	GACTTCTGGGAGGTGCACTGGGCGATCCCCACCCCGCGCGCTGAAGAAGAAGAAGAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
QY	988	GTGACCGTGTGACGTGGCGGACGCTACTTACGGTGCCTCTGGACGAGGACTTCCGC	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg	280
QY	1048	AAGTACACCCGCTTACCATCCCAGCATCAACAACGAGACCCCGCGCATCCCGCTACCAG	1107
Db	281	LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln	300
QY	1108	TACAACGTGTGCCCGCCAGGGCTGGAAGGGAGCGCCCGACATCTTCCAGAGCAGCATGACC	1167
Db	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320
QY	1168	AGATCTGTGAGCGCTTCCGGGCGCGAACCCCGAGATCGTGATCTACACAG-----GCC	1221
Db	321	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp	340
QY	1222	CCCTGTACTGTGGCAGCAGCTTGGAGATCGGCAGCACCGCGCAAGATCTGAGGAGCTG	1281
Db	341	AspLysThrValGlySerAspLeuGluIleGlyGlnHisArgThrIysIleGluGluLeu	360
QY	1282	CGCAAGCACCTGTGCTGGGCTTCCACACCCCGACAAGAAGCACCAAGAAGGAGCCC	1341
Db	361	ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro	380
QY	1342	CCCTTCTGTGGATGGGCTAGAGCTGCACCCCGACAAGTGGACCGTGGACGCCCATCGAG	1401
Db	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal	400
QY	1402	CTGCCCCAGAGGAGAGCTGCACCGTCAACGACATCCAGAGCTGTGTGGGCAAGCTGAAC	1461
Db	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
QY	1462	TGGGCCACGACAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGTCAAGCTGTCCGCGGC	1521
Db	421	TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly	440
QY	1522	GCCAAAGCCCTGACCGATCTGTGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAG	1581
Db	441	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu	460
QY	1582	AACCGCAGATCTCTGCGGAGCCCGTGCACGGCGGTGTACTACGACCCACGACGAGGACCTG	1641
Db	461	AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu	480
QY	1642	GTGCGCCAGATCCAGAAGACGAGGGCCACGACCGAGTGGACCTACCAGATCTTACCAGGAGCCC	1701

[illegible]



Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
Qy 1825 CCCAAGTTCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGGACCGACTCTGG 1884  
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTrp 560  
Qy 1885 CAGCGACCTCGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTTGAAGCTGTGG 1944  
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580  
Qy 1945 TACAGCTGAGAGAGGAGCCATCATCGGCGCGAGACCTTCACGTGAGCGCGCGCC 2004  
Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
Qy 2005 AACCGGAGACAGATCGGACAGCGCGCTAGTGACACCGGCGCGCGAGAGATC 2064  
Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
Qy 2065 GTGAGCTGACCGAGACCAACCAAGACCGAGCTGCAGCGCCATCCAGCTGGCCCTG 2124  
Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
Qy 2125 CAGGACAGCGGCGAGGTGAACATCGTACCGACGACGACGACGACGACGACGACG 2184  
Db 641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
Qy 2185 CAGCGCCAGCCGACAGAGCGAGCGAGCTGTGAACACGATCATCGACGAGCTGATC 2244  
Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680  
Qy 2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCGCGCCACCAAGGCGCATCGCGCAACGAG 2304  
Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysGlu 700  
Qy 2305 CAGATCGACAACTGTGTGACAGGCGATCCGCAAGGTGCTGTTCTGAGCGGATCGAT 2364  
Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 15  
ABR55489  
ID ABR55489 standard; protein; 1003 AA.  
AC ABR55489;  
XX  
DT 11-AUG-2003 (first entry)  
XX  
DE Amino acid sequence of a HIV pol protein.  
XX  
KW Polymorphic marker; host response; HIV; AIDS related virus; vaccine;  
KW hepatitis related virus; HCV; HBV; drug resistance; pol.  
XX  
OS Human immunodeficiency virus.  
XX  
PN WO2003035097-A1.  
XX  
XX 01-MAY-2003.  
PD  
XX 23-OCT-2002; 2002WO-AU001450.  
FF  
XX 23-OCT-2001; 2001AU-00008425.  
PR  
XX (BPIP-) EPIPOP PTY LTD.  
PA  
XX Mallal S;  
PI  
XX WPI; 2003-449231/42.  
DR  
XX  
XX  
XX  
PT Determining the influence of variation in host genes on the selection of  
PT microorganisms with protein substitutions, comprises typing individuals  
PT of a cohort infected with a microorganism for an intrinsic polymorphic  
PT marker.  
XX  
XX Claim 22; Page 88-89; 157pp; English.

XX The specification describes a method of determining the influence of  
CC variation in host genes on selection of microorganisms with protein  
CC substitutions. The method comprises typing all individuals of a  
CC population of patients infected with a microorganism for at least one  
CC selected intrinsic polymorphic marker involved in the host response to  
CC the presence of the microorganism. The method is useful for examining  
CC selective pressures confronting a wide range of organisms that exhibit  
CC pathogenic traits in a host, such as bacteria, fungi, mycobacterium,  
CC viruses and virus-like particles; for examining microorganisms that have  
CC adapted to evolve rapidly, including HIV and AIDS related viruses and the  
CC hepatitis related viruses such as HCV and HBV. The method is useful for  
CC designing a vaccine to prevent or delay the emergence of drug resistance  
CC in patients treated with a particular drug specific for a microorganism  
CC where the drug affects the replication of the microorganism at the  
CC nucleotide or amino acid level. The present sequence represents a HIV pol  
CC protein, which is expected to provide optimal cytotoxic T lymphocyte  
CC (CTL) induced therapeutic protection to the cohort examined in that study  
XX  
SQ Sequence 1003 AA;  
Alignment Scores:  
Pred. No.: 1,48e-193 Length: 1003  
Score: 3494.00 Matches: 654  
Percent Similarity: 95.70% Conservative: 36  
Best Local Similarity: 90.71% Mismatches: 25  
Query Match: 76.40% Indels: 6  
DB: 6 Gaps: 3  
US-09-610-313B-31 (1-2463) x ABR55489 (1-1003)  
Qy 220 TTCTTCCGCGAGACCTGGCTTCCCGGCAAGGCGCGCGAGTCCCGGAGAGCAG 279  
Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheSerGluGln 20  
Qy 280 AACCGCGCAACAGCCCGACCGAGCTGCAGTGGCGGCGCAC-----AACCCC 333  
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGluAspAsnSer 40  
Qy 334 CGCAGCGAGCG 387  
Db 41 ThrSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60  
Qy 388 CTGTGCGAGCG 447  
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeu 80  
Qy 448 GACACCG 507  
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100  
Qy 508 AAGATGATCG 567  
Db 101 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleIleIle 120  
Qy 568 GAGATCTGCGGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627  
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
Qy 628 ATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACCTTCCCATCAGCCCATC 687  
Db 141 IleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIle 160  
Qy 688 GAGACCGTCCCGTGAAGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
Qy 748 CTGACCGAGGAGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807  
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
Qy 808 AAGATCACAAGATCG 867

Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
Qy 868 AAGGACACCAAGTGGCGAAGCTGGTGAACCTTCGCGAGCTGAACAAGCGCACCCAG 927  
Db 221 LysAspSerThrLysTyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
Qy 928 GACTTCTGGAGGTGCAGCTGGGATCCCCACCCCGCGCGCTGAAGAAGAAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLysLysLysSer 260  
Qy 988 GTGACCGGTCTGGAGCTGGGCGAGCCCTACTTTCAGCGTGGCCCTGGACGAGCTTCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
Qy 1048 AAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACCCCGGCGATCCGCTACCCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
Qy 1108 TACAACGTGTCGCCGAGGCTGAAGGGGAGCCCGAGCATCTTCGAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
Qy 1168 AAGATCTCTGGAGCCTTCCGCGCGCGAACCCTCGAGATCGTGATCTACCCAG- ----GCC 1221  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
Qy 1222 CCCCTGTACGTGGCGAGCGACTCGGAGATCGCGAGACCCCGCGCATCGAGGAGCTG 1281  
Db 341 AspleuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
Qy 1282 CGAAGCACCTGTGCGCTGGGGCTTCACCAACCCCGACAAAGACCCAGAGAGAGCC 1341  
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrProAspLysLysHisGlnLysGluPro 380  
Qy 1342 CCCTTCTGTGGTGGCTACGAGCTGCACCCCGACAGTGGAGCGTGCAGCCCATCGAG 1401  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
Qy 1402 CTGCCCGAGAGAGAGCTGGACCGTGAACACATCCAGAGCTGGTGGGCAAGCTGAAC 1461  
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
Qy 1462 TGGGCCAGCCAGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTGTCTCGCGGC 1521  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
Qy 1522 GCCAAGCCCTGACCGACATCGTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCCGAG 1581  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
Qy 1582 AACCGGAGATCTCGCGAGCCGTCGACGCGCTGTACTACGACCCCGACGAGGAGCTG 1641  
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480  
Qy 1642 GTGGCCGAGATCCAGAGAGCGGCCACGACAGTGGACCTTACCAAGATCTACCGAGGCC 1701  
Db 481 IleAlaGluIleGlnLysGlnGlyGlnIleTrpThrTyrGlnIleTyrGlnGluPro 500  
Qy 1702 TTCAGAACCTGAAGACCGGCAAGTACGCCAGATGGCACCAGCCGACACCAAGAGCTG 1761  
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
Qy 1762 AAGCAGCTGACCGAGGCGGTGCAGAAGATGCCATGGAGAGCATCGTGATCTGGGGCAAG 1821  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
Qy 1822 ACCCCCAAGTTCGCCCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGAGCCGACTAC 1881  
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTyrTrpThrGluTyr 560  
Qy 1882 TGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1941  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy 1942 TGGTACCAGCTGGAGAGAGCCCATCATCGGCCCGAGACCTTTTACGTGAGCGGCC 2001  
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
Qy 2002 GCCAACCCGAGAGACCAAGATCGGCAAGCCCGGCTACGTGACGACCGCGCGCGAGAG 2061  
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620  
Qy 2062 ATCTGTAGCTGACCGAGAGACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCC 2121  
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
Qy 2122 CTGCAAGACAGCGCGAGCGAGGTGAACATCGTACCAGACAGCCAGTACGCTGGGCATC 2181  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
Qy 2182 ATCCAGGCCAGCCCGACAGAGCGAGCGAGCTGGTGAACACGATCATCGAGCTG 2241  
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
Qy 2242 ATCAAGAGAGAGAGGTGTACTCTGAGCTGGGTGCCGCCCAAGGGCATCGCGGCAAC 2301  
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700  
Qy 2302 GAGCAGATCGACAGCTGGTGCAGCAAGGCGCATCCGCAAGGTGTCTTCTCGACGCGCATC 2361  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
Qy 2362 GAT 2364  
Db 721 Asp 721

Search completed: June 2, 2005, 04:42:47  
Job time : 231.833 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 02:53:39 ; Search time 40.8333 Seconds  
(without alignments)  
11607.276 Million cell updates/sec

Title: US-09-610-313B-31  
Perfect score: 4573  
Sequence: 1 gtcagccaccatggccga.....gggtagcaccggtagaattc 2463

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgm2\_1/USPO.spool/US09610313/runat 31052005 155137 15159/app query.fasta\_1.7893  
-DB=PIR\_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=BITS -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09610313 @CGN 1 1 189 @runat 31052005 155137 15159 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: PIR\_79:\*  
2: PIR\_1:\*  
3: PIR\_2:\*  
4: PIR\_3:\*  
5: PIR\_4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3505.5	76.7	1002	2	pol polyprotein -
2	3497.5	76.5	1002	1	HIV-1 retropepsin
3	3492	76.4	1003	1	HIV-1 retropepsin
4	3475	76.0	1015	1	HIV-1 retropepsin
5	3453	75.5	1003	2	pol polyprotein -
6	3450	75.4	1003	1	HIV-1 retropepsin
7	3449	75.4	1003	1	HIV-1 retropepsin
8	3439	75.2	1012	1	HIV-1 retropepsin
9	3211	70.2	1027	1	HIV-1 retropepsin
10	3100	67.8	912	2	pol polyprotein -
11	3094	67.7	902	2	pol polyprotein -
12	2771	60.6	559	2	reverse transcript
13	2751	60.2	559	2	reverse transcript
14	2350	51.4	1039	2	pol polyprotein -

ALIGNMENTS

RESULT 1

S54378

pol polyprotein - human immunodeficiency virus type 1 (fragment)

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004

R;Accession: S54378

R;Theodore, T.; Buckler-White, A.J.

submitted to the EMBL Data Library, July 1989

A;Reference number: S54377

A;Accession: S54378

A;Status: preliminary

A;Molecule type: Genomic RNA

A;Residues: 1-1002 <THE>

A;Cross-references: UNIPROT:P12499; EMBL:M22639; NID:G329377; PIDN:AAA45366.1; PID:G3293

C;Superfamily: pol polyprotein

C;Keywords: polyprotein

Alignment Scores:

Pred. No.:	2.59e-155	Length:	1002
Score:	3505.50	Matches:	655
Percent Similarity:	95.69%	Conservative:	34
Best Local Similarity:	90.97%	Mismatches:	26
Query Match:	76.66%	Indels:	5
DB:	2	Gaps:	3

US-09-610-313B-31 (1-2463) x S54378 (1-1002)

Qy	220	TTCTTCGGCAGGACCTGGCTTCCCGCAGGAGCCCGCGAGTTCCCGCAGCGAGCAG 279
Db	1	PhapheargluaspLeuAlaPheProGlnGlysalaglyGluLeuSerSerGluGln 20
Qy	280	AACCGGCCAACAGCCCGCAGCGAGCTGAGGTGCGCGC---GACAAACCCCGC 336
Db	21	ThrArgAlaasnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40
Qy	337	AGCGAGCGCGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCTTC 390
Db	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu 60

```
QY 391 TGGCAGCGCCCTGGTGGAGCATCAAGGTGGGGCGGCAGATCAAGGAGGCGCTGTCTGGAC 450
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 TrpGlnArgProLeuValThrIleYsIleGlyGlnLeuYsGluAlaLeuLeuAsp 80
QY 451 ACCGGCGCGCAACACCGTGTGGAGGAGATGAGCTGCGCGGCAAGTGGAAAGCCCAAG 510
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100
QY 511 ATGATCGGGCGGATCGGCGCTTCATCAAGGTGCGCAGTACACAGATCCTGATCGAG 570
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 MetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu 120
QY 571 ATCTCGGCAAGAACGCAATCGGCACCGTGTCTGATCGGCCACCCCGTGAACATCATC 630
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140
QY 631 GGCGGCAACATGTCAGCCAGCTGGCTGCACCTGAACCTCCCATCAGCCCCATCGAG 690
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160
QY 691 ACCGTGCGGTGAAGTGAAGCCCGCATGAGCGGCGCCCAAGTGAAGCAGTGGCCCCCTG 750
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180
QY 751 ACCGAGGAGAAAGATCAAGGCCCTGACCGCATCTCGGAGGAGATGAGAAAGAGGCAAG 810
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200
QY 811 ATCACAAGATCGGCCCGGAGAACCCCTACAAACACCCCGCTGTCGCCATCAAGAAAG 870
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 IleSerArgValGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220
QY 871 GACAGACCAAGTGGCGAAGCTGTGGTGAATCCGGGAGCTGAACAAGCGCACCCAGGAC 930
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240
QY 931 TTCCTGGAGGTGCAGCTGGGCATCCCCCACCCCGCGCTGAAGAAGAAAGAGAGCGTG 990
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260
QY 991 ACCGTGCTGGACGTGGCGCAGCTACTTCAGCGTCCCGCTGACGAGGACTTCGCCAAG 1050
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLys 280
QY 1051 TACACGCGCTTCACCATCCCGACATCAACAAGAGACCCCGCGCATCCGCTACCGATAC 1110
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300
QY 1111 AACGTGTCGCCAGAGGCTGAAGGCGAGCCCGCAGCATCTCCAGAGCAGCATGACCAAG 1170
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320
QY 1171 ATCCTGGAGCGCTTCGCGCGCCGCAACCCCGAGATCGTGATCTACCAAG-----GCCCCC 1224
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340
QY 1225 CTGTAGTGGGAGGACCTGGAGATCGGCACGCGCCCAAGATCGAGGAGCTGGCG 1284
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360
QY 1285 AAGCAGCTGCTGGCTGGGCTTCACCAACCCCGCACAAGAAGCACAGAAAGAGGAGCCCCC 1344
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380
QY 1345 TTCCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG 1404
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400
QY 1405 CCGAGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGGCAGCTGAACCTGG 1464
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420
```

```
QY 1465 GCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTCGCGCGGCC 1524
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440
QY 1525 AAGGCCCTGACGACATCGTGCCTCGACCGAGGAGGCGCGAGCTGGAGCTGGCCGAGAAC 1584
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 LysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGluAsn 460
QY 1585 CCGAGATCCTCTCGCGAGCCCGTGCACGCGCTGTACTACGACCCCAAGGACCTGGTG 1644
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480
QY 1645 GCCAGATCCAGAAGCAGGCGCACACAGTGCACCTACAGATCTACAGAGAGCCCTTC 1704
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500
QY 1705 AAGAAGCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACACGACGCGAAG 1764
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520
QY 1765 CAGCTGACGAGCGCGTGCAGAAAGATCGCCATGAGAGCATCGTGATCTGGGGCAAGACC 1824
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
521 GlnLeuAlaGluValGlnLysIleSerThrGluSerIleValIleTrpGlyLysThr 540
QY 1825 CCCAAGTTCGCGCTGCCATCCAGAAGAGACCTGGAGACCTGTGTGGACCGCATCTGG 1884
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpValGluTyrTrp 560
QY 1885 CAGGCCACCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCTGTGTGAGCTGTGG 1944
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580
QY 1945 TACCAGCTGAGAAGGAGCCCATCATCGCGCGCAGACCTTCTACGTGCACGCGCGGCC 2004
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValaspGlyAlaAla 600
QY 2005 AACCGCGAGACCAAGATCGGCAAGCGCGCTACGTACCGACCGCGCGCGCGCAAGATC 2064
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620
QY 2065 GTGAGCCTGACCGAGACCAACCAAGACCGAGCTCGAGCCATCCAGCTGGCCCTG 2124
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
QY 2125 CAGGACAGCGCGCAGCGAGGTGAACATCGTGACCGACAGCGAGTACGCCCTGGGCATCATC 2184
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660
QY 2185 CAGGCCACCGCCGACAAGACGAGAGCGAGCTGGTGAACAGATCATTCAGAGCAGCTGATC 2244
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680
QY 2245 AAGAAGGAGAGGTGTACTCTGACCTGGTGGTGGCCCGCCACAGGGCATCGCGGCAACGAG 2304
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700
QY 2305 CAGATCGACAAGCTGTGTGAGCAGGCGCATCCCGCAAGTGTCTTCTGTGACGCGCATCGAT 2364
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 2
GNLJND
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: Host Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: JQ0067
R;Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.;
Gene 81, 275-284, 1989
A;Title: Nucleotide sequence of HIV-NDK: a highly cytopathic strain of the human immunoc
A;Reference number: JQ0065; MUID:90034200; PMID:2806917
A;Accession: JQ0067
```

A:Molecule type: DNA  
 A:Residues: 1-1002 <SPI>  
 A:Cross-references: UNIPROT:P18802; GB:M27323; NID:g328154; PIDN:AAA44869.1; PID:g328158  
 F:56-154/Product: retropepsin #status predicted <RTP>  
 C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
 C:Genetics:  
 A:Gene: pol  
 C:Superfamily: pol polypeptide  
 C:Keywords: ADS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle  
 F:56-154/Product: retropepsin #status predicted <RTP>  
 F:80/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
 Pred. No.: 6.08e-155 Length: 1002  
 Score: 3497.50 Matches: 655  
 Percent Similarity: 95.69% Conservative: 34  
 Best Local Similarity: 90.97% Mismatches: 26  
 Query Match: 76.48% Indels: 5  
 DB: 1 Gaps: 3

US-09-610-313B-31 (1-2463) x GNLJND (1-1002)

QY	220	TTCTTCGCGAGGACCTGGCCCTCCCGCAGGCAAGCGCGAGTTCCTCCCGAGCGAGCAG	279
DB	1	PhePheArgGluLeuAlaPheProGlnGlyLysAlaGlyGluPheSerSerGluGln	20
QY	280	AACCGCGCCAAACGCCCCACCGCGCGAGCTGCAGTG---CGCGCGCAACCCCGCGC	336
DB	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu	40
QY	337	AGCGAGCGCGCGCGAGCGCGAGCGCGACCCCTG-----AACTTCCCGAGATCACCCCTG	390
DB	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu	60
QY	391	TGGCAGCGCCCTCGTGAGCATCAAGTGGCGCGCGCAGATCAAGGAGGCGCTGCTGGAC	450
DB	61	TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp	80
QY	451	ACCGCGCGCGACACCGCTGCTGGAGGAGATGAGCTGCCCGCAAGTGGAAAGCCCAAG	510
DB	81	ThrGlyAlaAspThrValLeuGluGluLeuAsnLeuProGlyLysTrpLysProLys	100
QY	511	ATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGCAGTACGACGATCGCTGATCGAG	570
DB	101	MetIleGlyIleGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu	120
QY	571	ATCTGCGCGCAAGAGCGCATCGGCACCGTGTGATCGCGCGCCACCCCGTGAACATCATC	630
DB	121	IleCysGlyTyrlsAlaMetGlyThrValLeuValGlyProThrProValAsnIle	140
QY	631	GGCGCAACATGTGTACCCAGCTGGCTGCACCTGAACTTCCCATCAGCCCCCATCGAG	690
DB	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
QY	691	ACCGTGGCGTGAAGTGAAGCCCGCATGAGCGCGCCCGCAAGTGAAGCATGTCGCCCTG	750
DB	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
QY	751	ACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGCGAGGATGGAGAGGGGCAAG	810
DB	181	ThrGluGlnIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys	200
QY	811	ATCAACCAAGATCGCGCGCGAGAACCCCTTACAACACCCCGCTGTTCGCGCATCAAGAAAG	870
DB	201	IleSerArgIleGlyProGluAsnProTyrrAsnThrProIlePheAlaIleLysLys	220
QY	871	GACAGCACCAAGTGGCGAGCTGGTGACTTCGCGAGCTGAACAAGCGCACCCAGGAC	930
DB	221	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	240
QY	931	TTCTGGAGGTGCAGCTGGCGATCCCGCACCCCGCGCTGAGAGAGAGAGAGCGTG	990
DB	241	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	260

QY	991	ACCGTGTGTGAGTGGCGGAGCGCTACTTCAAGCTGCCCCCTGGAGGAGACTTCCGCAAG	1050
DB	261	ThrValLeuAspValGlyAspAlaTyrrPheSerValProLeuAspGluAspPheArgLys	280
QY	1051	TACACCGCTTACCATCCCGCAGCATCAACAACAGAGACCCCGCGCATCCGCTACCACTAC	1110
DB	281	TyrrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyLysIleArgTyrrGlnTyrr	300
QY	1111	AACGTGTGTGCGCGAGGCTGGAAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATCAAG	1170
DB	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
QY	1171	ATCTGTGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGATCTACCAAG-----GCCCCC	1224
DB	321	IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrrGlnTyrrMetAspAsp	340
QY	1225	CTGTACGTGGCGCAGCAGCTGGAGATCGGCGCAGCAGCGCGCAGATCGAGGAGCTGCGC	1284
DB	341	LeuTyrrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	360
QY	1285	AAGCACCTGTGCTGGCTGGGCTTCCACACCCCGCAGCAAGAACGACCAAGAGGAGCCCGCC	1344
DB	361	GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	380
QY	1345	TTCTGTGTGATGGCTACGAGCTGCACCCCGCAGCAGTGGACCGCTGCGCGCCATCGAGCTG	1404
DB	381	PheLeuTrpMetGlyTyrrGluLeuHisProAspLysTrpThrValGlnProIleAsnLeu	400
QY	1405	CCCGAAGAGAGAGCTGGACCGTGACGATCATCAGAGCTGTGCGGCGCAAGCTCAACTGG	1464
DB	401	ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	420
QY	1465	GCACGACGATCTACCCCGCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGGCC	1524
DB	421	AlaSerGlnIleTyrrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr	440
QY	1525	AAGCCCTGACCGACATCTGTCGCCCTGTACCGAGAGCGCGAGCTGGAGCTGGCGAGAAC	1584
DB	441	LysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuAlaGluAsn	460
QY	1585	CGCGAGATCTGCGCGAGCGCGTGCAGCGGTGTACTACACCCCGCAGCAAGGAGCTGTG	1644
DB	461	ArgGluIleLeuLysGluProValHisGlyValTyrrAspProSerLysAspLeuIle	480
QY	1645	GCAGAGATCCAGAAAGCAGCGCCACGACGATGGACCTTACAGATCTACAGGAGCCCTTC	1704
DB	481	AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrrGlnIleTyrrGlnProPhe	500
QY	1705	AAGAACCTGAAGACCGGCAAGTACCGCAAGATGGCACCGCCCGCACACCAAGAGCTGAAG	1764
DB	501	LysAsnLeuLysThrGlyLysTyrrAlaArgThrArgGlyAlaHisThrAsnAspValLys	520
QY	1765	CAGCTGACCGAGCGCTGCAAGAGATCGCCATCGAGAGCATCTGTGATCTGGGGCAAGACC	1824
DB	521	GlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr	540
QY	1825	CCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGAGCCGACTATGG	1884
DB	541	ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrrTrp	560
QY	1885	CAGCGCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGGTGAAGCTGTGG	1944
DB	561	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp	580
QY	1945	TACAGCTGAGAGAGAGCGCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC	2004
DB	581	TyrrGlnLeuGlnLysGluProIleIleGlyAlaGluThrPheTyrrValAspGlyValAla	600
QY	2005	AACCGCGAGACCAAGATCGGCAAGCGCGTACTGTACCGACCGCGCGCGCGCGAGAGATC	2064
DB	601	AsnArgGluThrLysLeuGlyLysAlaGlyTyrrValThrAspArgGlyArgGlnLysVal	620
QY	2065	GTGAGCGCTGACCGAGAGACCAACCAAGAGACCGAGCTGCGAGCGCCATCCAGTGGCCCTG	2124

```
Db      621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
|||:::|||||
QY      2125 CAGGACGCGGAGCGAGGTGAACATCGTACCGACGACGCGAGTACGGCTTGGGCATCATC 2184
|||||
Db      641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTrpAlaLeuGlyIleIle 660
|||||
QY      2185 CAGGCCAGCCGACAAAGACGAGAGCGAGCTGGTGAACAGATCATCAGACGAGCTGATC 2244
|||||
Db      661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680
|||||
QY      2245 AAGAAGAGAGAGTGTACCTGAGCTGGTGGTCCGCCACAAAGGCGATCGCGCGCAACGAG 2304
|||||
Db      681 LysLysGlnLysValTrpLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700
|||||
QY      2305 CAGATCGACAGCTGTGTGACCAAGGCGATCCGCAAGGTGTGTCTCGACGCGCATCGAT 2364
|||||
Db      701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
|||||
RESULT 3
GNVWL
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LAV-1a)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 03-Jun-2002
C;Accession: A03966
R;Main-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A;Title: Nucleotide sequence of the AIDS virus, LAV.
A;Reference number: A90866; MUID:85099333; PMID:2981635
A;Accession: A03966
A;Molecule type: DNA
A;Residues: 1-1003 <WAI>
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle
F;57-135/Product: retropepsin #status predicted <RFP>
F;81/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:
Pred. No.: 1,09e-154 Length: 1003
Score: 3492.00 Matches: 655
Percent Similarity: 95.42% Conservative: 33
Best Local Similarity: 90.85% Mismatches: 27
Query Match: 76.36% Indels: 6
DB: 1 Gaps: 3

US-09-610-313B-31 (1-2463) x GNVWL (1-1003)
QY      220 TTCTTCGGGAGGAGCTGGCTTCCCGGCAAGCGCCGGAGTTCCTCCAGCGAGCAG 279
|||||
Db      1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLeuAlaArgGluPheSerSerGluGln 20
|||||
QY      280 AACCGCGCCAAACGCCACCGCCGAGCTCGAGTG-----CGCGCGCAACACCCC 333
|||||
Db      21 ThrArgAlaAsnSerProThrArgGluLeuGlnValTrpGlyArgAsnAsnSer 40
|||||
QY      334 CGCAGCGCAGCGCGCGCGCAGCGCCAGCGCACCGCTG-----AACTTCCCGCAGATCAC 387
|||||
Db      41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
|||||
QY      388 CTGTGCGAGCCCGCTTGTGAGCATCAAGTGGCGCGCCAGATCAAGGAGGCCCTGCTG 447
|||||
Db      61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
|||||
QY      448 GACACCGCGCCGACGACCGCTGCTGAGGAGATGAGCTGCGCGCGCAAGTGGAGCCC 507
|||||
Db      81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100
|||||
QY      508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGAGTACGACCATCTCTGATC 567
|||||
```

```
101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIle 120
|||||
568 GAGATCTGGGCAAGAGGCCCATCGCACCGGTGCTGATCGGCCCCACCCCGTGGACATC 627
|||||
121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
|||||
628 ATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTCAACTTCCCATCAGCCCCATC 687
|||||
141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
|||||
688 GAGACCGTCCCGGTGAAGCTGAAGCCCGCATGAGCGCCCAAGGTGAAGCAAGTGGCCC 747
|||||
161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
|||||
748 CTGACCGGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAGGAGGCG 807
|||||
181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
|||||
808 AAGATCACCAAGATCGGCCCGGAGAACCCCTACACACCCCGCTGTCGCCCATCAAGAAG 867
|||||
201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
|||||
868 AAGGACAGCACCAAGTGGCGCAAGCTGTGGACTTCCCGGAGCTGAACAGCGCACCCAG 927
|||||
221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
|||||
928 GACTTCTGGAGGTGCGCTGGGCATCCCGCCACCCCGCGCTGCTGAAGAGAGAGAGAGC 987
|||||
241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
|||||
988 GTGACCGTCTCGACGCTGGCGCAGCGCTACTTTCAGGTGCTCCCTGACGAGACTTCCGC 1047
|||||
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280
|||||
1048 AAGTACACCGCTTACCATCCCGAGCATCAACACAGACACCCCGCGCATCCGCTACAG 1107
|||||
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
|||||
1108 TACACGCTCTGCCCGGCGCTGGAGGCGACCGCCCGCATCTTCCAGAGCAGCATGACC 1167
|||||
301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
|||||
1168 AAGATCTCGAGCCCTTCCGCGCCCGCAACCCGAGATCGTGTATCTACACAG-----GCC 1221
|||||
321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
|||||
1222 CCCCTGTACGTGGGACGACCTGGAGATCGCGCAGCACCGCGCCCAAGATCGAGGAGCTG 1281
|||||
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
|||||
1282 CGCAAGCACCTGCTGGCGTGGGGCTTCACCCCGCGACAGAACAGACACAGAGGAGGCC 1341
|||||
361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380
|||||
1342 CCCTTCTGTGATGGGCTACGAGCTGACCCCGACCCCGACAAGTGGACCGTGCACCATCGAG 1401
|||||
381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
|||||
1402 CTGCCCAGAGAGAGAGCTGGACCGGTGAACGACATCCAGAGCTGTGTGGCGCAAGCTGAAC 1461
|||||
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
|||||
1462 TGGGCCGAGCAGATCTTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGGC 1521
|||||
421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
|||||
1522 GCCAAGCGCTGACACATCGTGCCTCCCTGACCGAGAGCGCGAGCTGAGAGCTGGCGCAG 1581
|||||
441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluAlaGlu 460
|||||
1582 AACCGCGAGATCTTGGCGAGCGCGTGCACGGCGTGTA CTACGACCCCGAGGAGGACCTG 1641
|||||
```





```
Db 281 AlaTyr-PheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
|||
1072 AGCATCAACAGACAGACCCCGCATCCGCTACAGTACAACTGCTGCCACAGGCTGG 1131
|||
301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
|||
1132 AAGGCGAGCCCGACGATCTTCCAGAGCAGCATCACCAAGATCTCGAGCCCTTCCGCGCC 1191
|||
321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
|||
1192 CGCAACCCCGAGATCGTGATCTACCAAG-----GCCCGCCCTGTACGTGGCGACGACCTG 1245
|||
341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
|||
1246 GAGATCGGCCAGCACCGCGCCAGATCGAGGAGCTGGCAGCACCTGCTCGCTGGGCG 1305
|||
361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
|||
1306 TTCACACACCCCGACAGAGCACCAAGAGCGCCCTTCTGTGGATGGGCTACGAG 1365
|||
381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
|||
1366 CTCACCCCGACAAAGTGGACCGCTGCAGCCCATCGAGCTGCCCGAGAGAGAGCTCGAAC 1425
|||
401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
|||
1426 GTCAACGACATCCAGAGCTGGTGGCGACAGCTAACTGGCGCCAGCAGATCTACCCCGCC 1485
|||
421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
|||
1486 ATCAAGTGGCGCCAGCTGTCAAGCTGCTCGCGCGGCCAAGCGCCCTGACCGACATCGTG 1545
|||
441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
|||
1546 CCCTGACCGAGGAGCGCGAGCTGGAGCTGCCGAGAACCGCGAGATCTCGCGCGAGCCC 1605
|||
461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLysLysGluPro 480
|||
1606 GTCACGCGGTGTACTACGACCCCGACGAGGAGCTGGTGGCCGAGATCCAGAGCAGGCG 1665
|||
481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
|||
1666 CACGACGAGTGGACCTTACAGATCTTACAGAGCGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
|||
501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnLysGluProPheLysAsnLeuLysThrGlyLys 520
|||
1726 TACGCCAAGATGCGCACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAG 1785
|||
521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
|||
1786 AAGATCGCCATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATC 1845
|||
541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
|||
1846 CAGAAGGAGACCTGGGAGACCTGTGGACCGACTACTTGGCAGCGCCACCTGATCTCCCGAG 1905
|||
561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
|||
1906 TGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACAGCTGGGAAGAGAGGCC 1965
|||
581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuLysLysGluPro 600
|||
1966 ATCATCGGCGCAGACCTTCTAGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGC 2025
|||
601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
|||
2026 AAGCGCGCTACGTGACCGACCGCGCGCGCGCAGAGATCGTGAGCTTGACCGAGACACC 2085
|||
621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
|||
2086 AACCAGAAGACCGAGCTGCAGGCGCATTCAGCTGGCGCTGCAGGACAGCGCGAGCGGTG 2145
|||
```

```
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
|||
2146 AACATCGTCACCGACAGCCAGTACGCTGGGCATCATCAGCGCCAGCCCGACAGAGC 2205
|||
661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
|||
2206 GAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGAGAGTGTACCTG 2265
|||
681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
|||
2266 AGCTGGTCCCGCCCGCCACAGGCGCATCGCGCGACAGCAGATCGACAGCTGGTGAGC 2325
|||
701 AlaTrpValProAlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSer 720
|||
2326 AAGGGCATCCGCAAGTGTCTTCTCTGGACGCGCATCGAT 2364
|||
721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
|||
RESULT 5
T09440
pol polyprotein - human immunodeficiency virus type 1 (strain JRFL) (fragment)
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T09440
R:Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A:Reference number: Z16673
A:Accession: T09440
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1003 <PAN>
A:Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 7,03e-153 Length: 1003
Score: 3453.00 Matches: 646
Percent Similarity: 95.01% Conservative: 39
Best Local Similarity: 89.60% Mismatches: 30
Query Match: 75.51% Indels: 6
DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x T09440 (1-1003)
QY 220 TTCCTCCGCGAGACCTGCGCTTCCCGCAGGCGAAGCCCGGAGTTCCCGCAGGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCAAACAGCCCGCCAGCCGCGAGCTGCAGGTG-----CGCGCGCAACACCC 333
Db 21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTrpGlyArgAspSerAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCCAGGCGC-----ACCCTGAACCTTCCCGCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyAsnValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCGCTGTGTGACATCAAGTGTGGCGGCGCAGATCAAGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGACACCGCTGTGTGGAGGAGATGAGCTGCGCGCGCAAGTGAAGACCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGCGCCAGGTACGACAGATCTCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTGGCGCAGAGGCCCATCGGCACCGCTGCTGATCGGCGCCCGCCCGCTGACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
|||
```

QY 628 ATCGGCGCCAAATGCTGACCAAGCTGGGCTGCACCTGAACTTCCCATCAGCCCCATC 687  
Db 141 IleglyArgAenLeuThrGlnIleGlyCysThrLeuAenPheProIleSerProIle 160  
QY 688 GAGACCGTGGCCGCTGACCTGAAGCCCGGATGACGCGCCCAAGGTGAACAGTGGGCC 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
QY 748 CTGACCGAGGAGAAAGATCAAGCGCCCTGACCCCATCTGCGAGGAGATGGAGAGAGGC 807  
Db 181 LeuThrGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
QY 808 AAGATCAACCAAGATCGGCGCCAGAACCCCTACAAACCCCGTGTGCGCCATCAAGAAG 867  
Db 201 LysIleSerLysIleGlyProGluAenProLysAenThrProValPheAlaIleLysLys 220  
QY 868 AAGCAGCAGCAAGTGGCGCAAGCTGGTGACCTCCGCGAGCTGACACGCGCACCCAG 927  
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAenLysLysThrGln 240  
QY 928 GACTTCTGGAGGTGACCTGGGATCCCGCATCCCGCCCGGCTGAAAGAAAGAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysArgLysSer 260  
QY 988 GTGACCGTGTGGACGTGGCGCGCCCTACTTCAGCGTGCCTCGACGAGACTTCGCG 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspLysAspPheArg 280  
QY 1048 AAGTACACCGGCTTCACCATCCCGATCAACACGAGACCCCGGATCCGCTACCGAG 1107  
Db 281 LysTrpThrAlaPheThrIleProSerIleAenAenGluThrProGlyIleArgTrpGln 300  
QY 1108 TACAACGTGTGCCCCAGGGCTGAAGGGCAGCCCGACATCTTCCAGAGCAGATGACC 1167  
Db 301 TyrAenValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCGAG -GCC 1221  
Db 321 LysIleLeuGluProPheArgLysGlnAenProAspIleIleIleIleIleIleIleIle 340  
QY 1222 CCCCTGTACGTGGGACGACCTGGAGATCGGCAGACCCCGGCAAGATCGAGAGCTG 1281  
Db 341 AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeu 360  
QY 1282 CGCAAGCACCTGTGCTGGGGTTCACACCCCGCACAGAGAGAGAGAGAGAGAGAGAG 1341  
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1342 CCCTTCTGTGGATGGGCTACGAGCTGCACCCCGACCAAGTGGACCGTGCAGCCCATCGAG 1401  
Db 381 PropheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
QY 1402 CTGCGCGAGAGAGAGCTGACCGTGAACGACATCCGAAGCTGGTGGGCAAGCTGAAC 1461  
Db 401 LeuProGluLysAspSerTrpThrValAenAspIleGlnLysLeuValGlyLysLeuAen 420  
QY 1462 TGGCGCAGCCAGATCTACCCCGCATCAAGTGGCGCAGCTGCAAGCTGCTCGCGGCG 1521  
Db 421 TrpAlaSerGlnIleTrpAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440  
QY 1522 GCCAAGCGCCCTGACCGACATCGTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAG 1581  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
QY 1582 AACCGGAGATCTCGCGGAGCCGTCGACCGGCTGTACTAGACCCCGAGCAAGAGACTG 1641  
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTrpTrpAspProSerLysAspLeu 480  
QY 1642 GTGCGCGAGATCCAGAGAGGCGCCAGCACAGTGGACCTACCAAGATCTACCGAGCC 1701  
Db 481 IleAlaGluLeuGlnLysGlnGlyGlnGlyGlnTrpThrGlnIleTrpThrGlnGluPro 500

QY 1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGCCACCGCCACACCAACGACGTG 1761  
Db 501 PheLysIleLeuLysThrGlyLysTrpAlaArgThrArgGlyAlaHisThrAsnAspVal 520  
QY 1762 AAGCAGCTGACCGAGGCGGTGCAAGATGCCATGCCATGAGAGCATCTGTGATCTGGGGCAAG 1821  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAenGlnSerIleValIleTrpGlyLys 540  
QY 1822 ACCCCCAAGTTCGCCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACTAC 1881  
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTrp 560  
QY 1882 TGGCAGCGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1941  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAenThrProProLeuValLysLeu 580  
QY 1942 TGTATACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCC 2001  
Db 581 TrpTrpGlnLeuGluLysGluProIleValGlyAlaGluThrPheTrpValAspGlyAla 600  
QY 2002 GCCAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGGCGCGCAGAG 2061  
Db 601 AlaAenArgGluThrLysLeuGlyLysAlaGlyTrpValThrAenArgGlyArgGlnLys 620  
QY 2062 ATCTGTAGCTGACCGAGACCAACACAGAGACCGAGCTGAGGAGCATCCAGCTGGGCC 2121  
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
QY 2122 CTGCAGGACGCGCAGCGAGGTGAACATCGTGACCGACGACGAGTACGCGCTCGGCATC 2181  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTrpAlaLeuGlyIle 660  
QY 2182 ATCCAGGCGCCAGCCGCAAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTG 2241  
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
QY 2242 ATCAAGAGGAGAGGTGTACTGTAGCTGGGTGGTCCCGCCACAGAGGAGCATCGGGGCAAC 2301  
Db 681 IleLysLysGluLysValTrpLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700  
QY 2302 GAGCAGATCCAGCAGCTGTGTGAGCAAGGGCATCCGCAAGGTGCTGCTTCGACGCGCATC 2361  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
QY 2362 GAT 2364  
Db 721 Asp 721

## RESULT 6

B44001

HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (strain YU-2)  
N;Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly.  
C;Species: human immunodeficiency virus type 1, HIV-1

A;Note: host Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004

C;Accession: B44001

R;Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.

J. Virol. 66, 6587-6600, 1992

A;Title: Complete nucleotide sequence, genome organization, and biological properties of

A;Reference number: A44001; MUID:93021387; PMID:1404605

A;Accession: B44001

A;Molecule type: DNA

A;Residues: 1-1003 &lt;LI&gt;

A;Cross-references: UNIPROT:P35963; GB:M93258

C;Comment: This protein is synthesized as a gag-pol polyprotein.

C;Comment: Specific enzymatic cleavages may yield mature proteins including proteinase,

C;Genetics:

A;Gene: pol

C;Superfamily: pol polyprotein

C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo

F;57-155/Product: retropepsin #status predicted &lt;RTP&gt;

F;81/Active site: Asp (shared with dimeric partner) #status predicted

## Alignment Scores:

Pred. No.: 9.68e-153 Length: 1003  
Score: 3450.00 Matches: 647  
Percent Similarity: 94.87% Conservative: 37  
Best Local Similarity: 89.74% Mismatches: 31  
Query Match: 75.44% Indels: 6  
DB: 1 Gaps: 3

US-09-610-313B-31 (1-2463) x B44001 (1-1003)

```
QY 220 TTCTTCGCGAGACCTGGCTTCCCTCCCGAGGCAAGCCCGCGAGTTCCCGAGGACAG 279
Db 1 PhePheArgGluAsePLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
QY 280 AACCGCGCCACACGCCCCACCAGCCGCGAGCTCAGGTG-----CGCGCGACACCCC 333
Db 21 ThrArgAlaAsnSerProIleAArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCCGCGCGCGCGAGCGCCAGGCAACCTCG-----AACTTCCCTCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGCCGCTGTGTGAGATCAAGTGGGGCCAGATCAAGAGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACGGCGCCGACGACCGCTGCTGGAGGAGATGAGCTGCCGCGAAGTGAAGGCC 507
Db 81 AspThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGGCGGCGAGTACACACAGATCCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProIle 120
QY 568 GAGATCTGCGGCAAGAGCCATCGGCACCGCTGATCGCGCCCGCCACCCCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGGCGGCACATGCTGACCCGAGCTGGGCTGCACCTGACCTGAACCTCCCATCAGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGTGAAGTGAAGCCCGCATGGAGCGGCCCAAGGTGAAGCAGTGGCCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGCGAGGAGATGGAAGAAGAGGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCCCGGAGACCCCTTACACACCCCGCTGTTCGCCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACACACCAAGTGGCCCAAGCTGGTGGACTTCGCGAGCTGAACCAAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGAGGTGACGTGGGCATCCCGACCCCGCGCTGGAAGAGAAGAAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGACGTGGCGGCGCTACTTTCAGGTGCGCCCTGGACGAGGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuHisGluAspPheArg 280
QY 1048 AAGTACACCGCTTCACCATCCCGAGCATCAACAAAGAGACCCCGCGCATCCGCTACCCAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyThrArgTyrGln 300
QY 1108 TACAAGTGTGCCCCAGGCGCTGGAAGGCGAGCCCGACGATCTTCCAGACGACATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
```

```
QY 1168 AAGATCTGTGAGCCCTTTCGCGCCCGCAACCCCGAGATCGTGTATCTACCAG-----GCC 1221
Db 321 ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTrpGlnTrpMetAsp 340
QY 1222 CCCTGTAGCTGGGCGACGACCTGGAGATCGGCGACGACCGCGCCCAAGATCGAGGAGCTG 1281
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 CGCAAGCACCTGTGCGCTGGGCTTCACACCCCGCAAGAACAGACACAGAGAGAGCC 1341
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCCTTCTGTGATGGCTACGAGCTGACCCGCAAGTGGACCTGCGAGCCCATCGAG 1401
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1402 CTGCCCGAAGAGAGAGCTGGACCGCTGAACGACATCCAGAAGCTGGTGGCGAAGCTGAAC 1461
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGCCGAGCCAGATCTACCCCGCATCAAGTTCGCGCAGCTGTGCAAGCTGCTGCCGCGC 1521
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1522 GCCAAGGCGCTGACCGACATCGTCCCTGACCGAGGCGGAGCTGCGAGCTGGCGCGAG 1581
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuAlaGlu 460
QY 1582 AACCGCGAGATCCTGCGGAGCCCGCTGACGCGGTGTACTACGACCCCGAGCAAGCACTG 1641
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480
QY 1642 GTGGCCGAGATCCAGAGAGCGGCCACGACGAGTGCACCTACCATCTACCGAGAGCC 1701
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
QY 1702 TTCAGAACTGAAGACCGGCAAGTACGCCAAGATCGCCACCGCCACACCAACGACGCTG 1761
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520
QY 1762 AAGCAGCTGACCGAGGCGCTGCAAGAGATCGCCATGAGAGAGCATCTGTATCTGGGCAAG 1821
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540
QY 1822 ACCCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGAGACCTGGTGGCCGAGCTAC 1881
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
QY 1882 TGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCTGGTGAAGCTG 1941
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1942 TGTATCAGCTGGAGAAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGCGCC 2001
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCAGGAGACCAAGATCGGCAAGCGCGCTAGCTACCGACCGGCGCGCGGAGAG 2061
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnLysGlyArgGlnLys 620
QY 2062 ATCGTGGAGCTGACCGAGACCAACACAGAGAGCCGAGCTGCAGGCCCATCCAGCTGGCC 2121
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
QY 2122 CTGCAGGACAGCGCGCAGGAGTGAACATCGTGACCCAGCAGCAGTACGCCCTGGGCATC 2181
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2182 ATCCAGGCGCGCGCAAGAGCGGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTG 2241
Db 661 IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
```

Qy 2242 ATCAAGAGGAGGAGGTGTACTGAGCTGGTGGCCGCCACAAAGGCATCGCGGCGAAC 2301

Db 681 ILeuLysGluLysValTyLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700

Qy 2302 GAGCAGATCGACAGCTGGTGAGCAAGGGCATCGCAAGGTGCTGTTCCTGGACGGCATC 2361

Db 701 GluGlnValAspLysLeuValSerAlaGlyLeuArgLysValLeuPheLeuAspGlyIle 720

Qy 2362 GAT 2364

Db 721 Asp 721

RESULT 7

GNVWA2

HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate ARV-2)

N;Antigens: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA pol

C;Species: human immunodeficiency virus type 1, HIV-1

A;Note: host Homo sapiens (man)

C;Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004

C;Accession: A03968

R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh

Science 227, 484-492, 1985

A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).

A;Reference number: A04003; MUID:85090453; PMID:2578227

A;Accession: A03968

A;Molecule type: DNA

A;Residues: 1-1003 <SAS>

A;Cross-references: UNIPROT:P03369; GB:K02007; NID:g328658; PIDN:AAB59876.1; PID:g328662

C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re

C;Genetics:

A;Gene: pol

C;Superfamily: pol polyprotein

C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle

F;57-147/Product: retropepsin #status: predicted <RTP>

F;81/Active site: Asp (shared with dimeric partner) #status: predicted

Alignment Scores:

Pred. No.: 1,08e-152 Length: 1003

Score: 3449.00 Matches: 646

Percent Similarity: 94.87% Conservative: 38

Best Local Similarity: 89.60% Mismatches: 31

Query Match: 75.42% Indels: 6

DB: 1 Gaps: 3

US-09-610-313B-31 (1-2463) x GNVWA2 (1-1003)

Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGGCAAGCGCCGCGAGTTCGCCAGGAGCAG 279

Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLeuAlaArgGluPheSerGluGln 20

Qy 280 AACCGCCCAACAGCCCCACAGCGCGAGCTGCGGTGCGCGC-----GACAACCCC 333

Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40

Qy 334 CGCAGCGAGCGCGCGCGCGAGCGCGAGCGCGAGCGCGCTG-----AACTTCCCCCAGATCACC 387

Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60

Qy 388 CTGTGGAGCGCGCCCTGGTGTGAGCATCAAGTGGGGCGCGAGATCAAGAGGCGCTCTGCTG 447

Db 61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyLeuLysGluAlaLeuLeu 80

Qy 448 GACACGGCGCGCGAGCACCGTGTGGAGGAGATGAGCTGCGCGCAAGTCGAAGCCC 507

Db 81 AspThrGlyAlaAspThrValLeuGluGluWetAsnLeuProGlyLysTrpLysPro 100

Qy 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGAGTACGACACGATCTCTGATC 567

Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProVal 120

Qy 568 GAGATCTGCGCGCAAGAGGCGCATCGGCACCGTGTGATCGCGGCCACCCCGGTGAACATC 627

Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

QY	1702	TTCAAGAACCTGAGACCGCGCAAGTAGCGCCCAAGATCGCACCGCCCAACACGACGTG	1761
Db	501	PhelyAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal	520
QY	1762	AAGAGCTGACCGAGCGCGTGAGAGATCGCCATGAGAGATCGTGATCTGGGCAAG	1821
Db	521	LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys	540
QY	1822	ACCCCAAGTTCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTAC	1881
Db	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpMetGluTrp	560
QY	1882	TGCGAGCGCACCTCGATCCCGAGTGGAGTTCTGTGAACACCCCGCCCTGTGTGAAGCTG	1941
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu	580
QY	1942	TGTTACAGCTGAGAGAGGCCATCATCTCGCGCGGAGACCTTCTACGTGGACGGGCC	2001
Db	581	TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla	600
QY	2002	GCCAAACCGGAGACCAAGATCGCGCAAGGCGGCTACGTGACCGACCGCGCGCAGAG	2061
Db	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys	620
QY	2062	ATCGTGAAGCTGACCGAGACCAACCAAGACCGAGCTGAGAGGCCATCCAGCTGGCC	2121
Db	621	ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla	640
QY	2122	CTCGAGACAGCGGCGAGGTGAACATCGTGACCGACGACGACGACGCGCTGGGCATC	2181
Db	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle	660
QY	2182	ATCCAGCGCCAGCCGACAGAGCGAGAGCGAGCTGGTGAAACAGATCATCGAGCAGCTG	2241
Db	661	IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu	680
QY	2242	ATCAAGAAGAGAGGTGTACCTGAGCTGGGTGCGCGCCGCCACCAAGGGCATCGCGCAAC	2301
Db	681	IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn	700
QY	2302	GAGCAGATGACAGCTGTGTGAGCAAGGGCATCGCAAGGTGCTGTTCGTGGACGGCATC	2361
Db	701	GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle	720
QY	2362	GAT 2364	
Db	721	Asp 721	
RESULT 8			
GNVWVL			
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly			
C;Species: human immunodeficiency virus type 1, HIV-1			
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004			
C;Accession: A03967			
R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.			
Nature 313, 450-458, 1985			
A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi			
A;Reference number: A93355; PMID:85111157; PMID:2982104			
A;Molecule type: DNA			
A;Accession: A03967			
A;Cross-references: UNIPROT.P03368			
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re			
C;Genetics:			
A;Gene: pol			
C;Superfamily: pol polyprotein			
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle			
F;66-164/Product: retropepsin #status predicted <RTP>			
F;90/Active site: Asp (shared with dimeric partner) #status experimental			
Alignment Scores:			
Pred. No.: 3.14e-152 Length: 1012			

Score:	3439.00	Matches:	647
Percent Similarity:	93.67%	Conservative:	34
Best Local Similarity:	89.00%	Mismatches:	28
Query Match:	75.20%	Indels:	18
DB:	1	Gaps:	4
US-09-610-313B-31 (1-2463) x GNVWVL (1-1012)			
QY	238	GCCTTCCCTCCAGGGCAAGCGCCGCGAGTTC	267
Db	4	AlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerPro	23
QY	268	-----CCGAGCGAGCAAGACCGCCCAACAGCCCAACAGCCCAAGCGCGAGTGCAGGTG---	318
Db	24	ThrIleSerSerGluGlnThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrp	43
QY	319	----CGGGCGGACACCCCGCGAGGCGGGCGCGAGCGCGCGAGCGCGCGACCTTG-----	369
Db	44	GlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPhe	63
QY	370	AATCTTCCCGACATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGCGCGCCAG	429
Db	64	AsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGlyGln	83
QY	430	ATCAAGAGGCGCTGTGTGACACCGCGCGCGACACACCGCTGTGTGAGAGATGACCTG	489
Db	84	LeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeu	103
QY	490	CCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTTCATCAAGGTGCCCGAG	549
Db	104	ProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGln	123
QY	550	TACGACCATCTGTATCGAGATCTCGCGCAAGAGCGCATCGGCACCGTGTCTGATCGGC	609
Db	124	TyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIleGlyThrValLeuValGly	143
QY	610	CCACACCCCGTGAACATCATCGCGCGCAACATCTGTCACCGAGCTGGGTGTCACCTGAAC	669
Db	144	ProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsn	163
QY	670	TTCCTCATCAGCCCATCGAGACCGTGGCGGTGAAGCTGAAGCCCGCATGAGCGGCC	729
Db	164	PheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyPro	183
QY	730	AAGTGAACAGTGGCCCTCGACCGAGGAGAAGATCAAGCCCTCGACCGCATCTCGAG	789
Db	184	LysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThr	203
QY	790	GAGATGGAGAAGGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTACACACCC	849
Db	204	GluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrPro	223
QY	850	GTGTTCCCATCAAGAGAAGACACACCAAGTGGCGCAAGCTGTGGAGCTTCCGCGAG	909
Db	224	ValPheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGlu	243
QY	910	CTGAACAAGCGACCCAGGACTTCTCGGAGGTGCAGCTGGGCATCCCCACACCCCGCGGC	969
Db	244	LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGly	263
QY	970	CTGAAGAAGAAGAGCGTGCCTGTGTGGAGCTGGCGGACGCGCTACTTTCAGCTGCC	1029
Db	264	LeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValPro	283
QY	1030	CTGGACGAGACTTTCGCAAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACC	1089
Db	284	LeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnGluThr	303
QY	1090	CCCGGATCCGCTTACCAGTACCAACGTGTGCGCCAGGCGTGGGAAGCGGACGCCAGCATC	1149
Db	304	ProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIle	323
QY	1150	TTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCGTG	1209

Db 324 PheGlnSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343  
Qy 1210 ATCTACCAG-----GCCCCCTGTGTACGTGGCGGACGACCTGGAGATCGGCCACGACCGC 1263  
Db 344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363  
Qy 1264 GCCAAGATCGAGGAGCTGGCGAAGCAGCCTGTGTGGTGGGCTTCACACCCCGACAG 1323  
Db 364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLys 383  
Qy 1324 AAGCACCAGAGAGCCCTCTCTGTGGATGGCTACGAGCTGCACCCCGACAAGTGG 1383  
Db 384 LysHisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrp 403  
Qy 1384 ACCGTGAGCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAAG 1443  
Db 404 ThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLys 423  
Qy 1444 CTGTGGGCAAGCTGAATGGGCCAGCAGATCTACCCCGGCATCAAGTGGCGCAGCTG 1503  
Db 424 LeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443  
Qy 1504 TGCAAGCTGTGGCGCGCCCAAGCCCTGACCCGATCGTGGCCCTGACCGAGAGGCC 1563  
Db 444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463  
Qy 1564 GAGCTGAGCTGGCGGAGAACCGGAGATCTCGCGGAGCCGCGACGGCGTGTACTAC 1623  
Db 464 GluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyr 483  
Qy 1624 GACCCAGCAGAGACCTGTGGCGCAGATCCAGAACGAGGCGCACGACCAAGTGAACCTAC 1683  
Db 484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTyrTrpTyrTyr 503  
Qy 1684 CAGATCTACAGGAGCCCTTCAAGAACCTGGAGACCGGACGATCGGACCAAGTGGCAC 1743  
Db 504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523  
Qy 1744 GCCCACAACCAAGCAGTGAAGCAGCTGACCGAGCCGCTGCAGAGAAGATCGCCATCGAGC 1803  
Db 524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543  
Qy 1804 ATCGTGTCTGGGCAAGACCCCAAGTTCGCTGCCCTCCATCCAGAAGGACCTGGAG 1863  
Db 544 IleValIleTyrGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTrpGlu 563  
Qy 1864 ACCTGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGGATTCGTGAACACC 1923  
Db 564 ThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThr 583  
Qy 1924 CCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACC 1983  
Db 584 ProProLeuValLysLeuTyrTrpTyrGlnLeuGlnLysGluProIleValGlyAlaGluThr 603  
Qy 1984 TTCTACGTGGACGGCGCCGCAACCGCGAGACCAAGATCGCAAGCGCGGTACGTGACC 2043  
Db 604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGlyLysAlaGlyTyrValThr 623  
Qy 2044 GACCGGGCGCGAGAGATCGTGAAGCTGACCGAGACCCAGACCCACCAAGAGACCGAGCTG 2103  
Db 624 AsnLysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643  
Qy 2104 CAGGCCATCCAGCTGGCGCTGCAGACAGCGGCGAGGAGTGAACATCGTGACCGACAGC 2163  
Db 644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSer 663  
Qy 2164 CAGTACCGCTGGGCATCATCCAGGCCCGACCGCGACAGACGAGCGAGCTGGTGAAC 2223  
Db 664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsn 683  
Qy 2224 CAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGTGGCGGCCAC 2283

Db 684 GlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProIleHis 703  
Qy 2284 AAGCGCATCGCGCGCAACGACGACATCGACAAGCTGTGTGACCAAGGGCATCGCAAGTGTG 2343  
Db 704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723  
Qy 2344 CTGTCTCTCGACGCGCATCGAT 2364  
Db 724 LeuPheLeuAspGlyIleAsp 730  
RESULT 9  
GNLJSI  
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz  
N;Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
C;Species: simian immunodeficiency virus SIVcpz  
A;Note: host Pan troglodytes (chimpanzee)  
C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C;Accession: S09984  
R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.  
Nature 345, 356-359, 1990  
A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.  
A;Reference number: S09983; MUID:90259077; PMID:2188136  
A;Accession: S09984  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1027 <HUE>  
A;Cross-references: UNIPROT:P17283; EMBL:X52154  
F;81-180/Product: retropepsin #status predicted <RTP>  
F;105/Active site: Asp (shared with dimeric partner) #status predicted  
C;Genetics:  
A;Gene: pol  
C;Superfamily: pol polypeptide  
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo  
F;81-180/Product: retropepsin #status predicted <RTP>  
F;105/Active site: Asp (shared with dimeric partner) #status predicted  
Alignment Scores:  
Pred. No.: 1,17e-141 Length: 1027  
Score: 3211.00 Matches: 595  
Percent Similarity: 88.92% Conservative: 63  
Best Local Similarity: 80.41% Mismatches: 77  
Query Match: 70.22% Indels: 5  
DB: 1 Gaps: 2  
US-09-610-313B-31 (1-2463) x GNLJSI (1-1027)  
Qy 159 CTGTGTGAAGTGGCGCAAGGAGGCGCACCATGATGAAGACTGCACCGAGCGCCAGGCCAAC 218  
Db 7 LeuLeuAlaValTrpAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26  
Qy 219 CTTTCTTCGCGAGGACCTGGCTTCCCGGAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 278  
Db 27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGluG1 46  
Qy 279 GAACCGCGCCACACGCGCCACCG 332  
Db 46 nAsnArgThrAsnGlyProThrAspArgGluLeuTrpValProGlyGlyArgGluGluPr 66  
Qy 333 CCGCAGGAGCG 392  
Db 66 oGlyGluLargGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86  
Qy 393 GCAGCGCCCTCTGTGAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452  
Db 86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106  
Qy 453 CGCGCGCGCGCACACCGTGTGTGGAGGAGATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 512  
Db 106 rGlyAlaAspThrValIleGluArgIleGlnLeuGlnGlyLeuTrpLysProLysMe 126  
Qy 513 GATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 572  
Db 126 tIleGlyGlyIleGlyPheIleLysValLysGlnPheAsnValHisIleGluI1 146





```
QY 481 ATGAGCTGCTCCGCGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTCATCAAG 540
Db 1 MetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 20
QY 541 GTGCGCCAGTACGACAGATCCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGCACCGTG 600
Db 21 ValLysGlnTrpAspGlnIleAlaIleGluLeuCysGlyHisLysAlaIleGlyThrVal 40
QY 601 CTGATCGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGC 660
Db 41 LeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuLeuThrGlnLeuGlyCys 60
QY 661 ACCCTGAACCTCCCATCGACCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGCGATG 720
Db 61 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 80
QY 721 GACGGCCCCAAGTGAAGAGTGCCTCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC 780
Db 81 AspGlyProLysValLysGlnTrpProLeuThrGlnLysIleLysAlaLeuIleGlu 100
QY 781 ATCTGCGAGGAGATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTAC 840
Db 101 IleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyr 120
QY 841 AACACCCCGCTGTTCGCCATCAAGAGAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 121 AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp 140
QY 901 TTCGGCGAGCTGAACAGCCACCCAGGACTTCTGGGAGGTGCAGCTGGCGCATCCCCAC 960
Db 141 PheArgGlnLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 160
QY 961 CCCGCGCGCTGAAGAGAAGAGAGCGTCAACCGTCTGGACGTGGCGGACGCTACTTTC 1020
Db 161 ProGlyGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 180
QY 1021 AGCGTGCCCTTGACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGATCAAC 1080
Db 181 SerValProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsn 200
QY 1081 AACGAGACCCCGCGCATCCCTACCACTACAGTGTCTGCCCGCCGCGGCTGGAGGGCAGC 1140
Db 201 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 220
QY 1141 CCAGAGATCTTCAGAGCAGATGACCAAGATCTCGGAGCGCTTCCGCGCGCCGCAACCCC 1200
Db 221 ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro 240
QY 1201 GAGATCGTGATCTACCAAG-----GCCCGCTGTACGTGGCGACGACCTGGAGATCGGC 1254
Db 241 GluIleIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 260
QY 1255 CAGCAGCGCCCAAGATCGAGGAGCTCGGCAAGCAGCTGTGCGTGGCGGCTTCACCAAC 1314
Db 261 GlnHisArgThrLysIleGluLeuArgArgHisLeuLeuArgTrpGlyPheThrThr 280
QY 1315 CCCGACAAGAGCAGCAGAGGAGCCCGCTCTCTGTGGATGGCTACGAGCTGCACCCC 1374
Db 281 ProAspLysLysHisGlnLysGluProPheLeuTrpIleGlyTyrGluLeuHisPro 300
QY 1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGAC 1434
Db 301 AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp 320
QY 1435 ATCCAGAGCTGGTGGCGCAAGTGAACCTGGCGCCAGCATCTACCCCGCGCATCAAGGTG 1494
Db 321 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal 340
QY 1495 CGCCAGCTGTGAAGCTGTGCGCGCGCCCAAGSCCTGACCGACATCGTGGCCCTGAC 1554
Db 341 ArgGlnLeuCysArgLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr 360
QY 1555 GAGGAGCGCGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGC 1614
```

```
Db 361 LysGluAlaGluLeuGluLeuAlaGluAsnArgGluLeuLeuLysThrProValHisGly 380
QY 1615 GTGTACTTACGACCCAGCAAGGACCTGTGTGCCGAGATCCAGAAGACGAGGGCCACGACCAG 1674
Db 381 ValTyrTyrAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyLeuGlyGln 400
QY 1675 TGGACCTTACAGATCTTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
Db 401 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 420
QY 1735 ATGGGACCGCCGACCAAGAGCTGAAGAGCTGACCGAGCGCTGCGAGAGATCGCC 1794
Db 421 MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 440
QY 1795 ATGAGAGATCGTGTCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGAG 1854
Db 441 ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu 460
QY 1855 ACCTGGAGACCTGTGTGGACCGACTACTTGGCAGCCACCTGGATCCCCGAGTGGAGTTC 1914
Db 461 ThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 480
QY 1915 GTGAACACCCCGCTGTGTGAAGCTGTGTACCGCTGGAGAGGAGGCCCATCATCGGC 1974
Db 481 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 500
QY 1975 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGC 2034
Db 501 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 520
QY 2035 TACGTGACCGACCGCGCGCGCAGAGATCGTGCAGCTGCAGCCGACGACCCACCAAG 2094
Db 521 TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrThrAsnGlnLys 540
QY 2095 ACCGAGCTGACGCGCATCTGAGCTGGCGCTCGAGACAGCGCGCAGGAGTGAACATCGTG 2154
Db 541 ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal 560
QY 2155 ACCGACAGCCAGTACGCGCTGGCGCATCATCCAGCGCCGCGCCGACAGAGCGAGCGAG 2214
Db 561 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 580
QY 2215 CTGCTGAACAGATCATCTGAGCAGCTGATCAAGAGAGAGAGTGTATCTGAGCTGGGTG 2274
Db 581 LeuValSerGlnIleIleGluHisLeuIleLysLysGluLysValTyrLeuAlaTrpVal 600
QY 2275 CCCGCCACAGAGGCGATCGCGCGCAGCAGCAGATCGACAGCTGGTGGAGCAAGGGCATC 2334
Db 601 ProAlaHisLysGlyIleGlyGlyAsnGluValAspLysLeuValSerAlaGlyIle 620
QY 2335 CGCAAGGTGTCTCTCTGGACGCGCATCGAT 2364
Db 621 ArgLysValLeuLeuAspGlyIleAsp 630

RESULT 11
T01668
pol polyprotein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T01668
R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.
Cell 46, 63-74, 1986
A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol
A;Reference number: Z14389; MUID:86245056; PMID:2424612
A;Accession: T01668
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-902 <Al>
A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28012.1; PID:g60230
C;Superfamily: pol polyprotein

Alignment Scores:
```

Pred. No.:	3,05e-136	Length:	902
Score:	3094.00	Matches:	574
Percent Similarity:	97.10%	Conservative:	28
Best Local Similarity:	92.58%	Mismatches:	16
Query Match:	67.66%	Indels:	2
DB:	2	Gaps:	1
US-09-610-313B-31 (1-2463) x T01668 (1-902)			
QY	511	ATGATCGGGGATCGGCGCTTCATCAAGGTGGCGCAGTACGACCAGATCCTGATCGAG	570
Db	1	MetilegilyleGlyPheIleYsValArgInTyAspGlnIleLeuIleGlu	20
QY	571	ATCTGGGCAAGAAGCCATCGGCACCGTGTGATCGGCCCCACCCCGTGAACATCATC	630
Db	21	IleCysGilylsalaleGlyThrIleLeuValGlyProThrProValAsnIleIle	40
QY	631	GGCGCAACATGTGACCCAGCTGGGTGCGACCTGAATCTCCCATAGCCCCATCGAG	690
Db	41	GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	60
QY	691	ACCGTCCCGTGAAGCTGAAGCCGCGCATGGACGGCCCAAGGTGAAGCAGTGGCCCTG	750
Db	61	ThrValProValYsLeuYsProGlyMetAspGlyProArgValYsGlnTrpProLeu	80
QY	751	ACCGAGGAGAAGATCAAGGCCCTGACCGGCATCTGCGAGGATGAGAGAGGGCAAG	810
Db	81	ThrGluGluYsIleYsAlaIleuThrGluIleCysLeYsAspMetGluYsGluYs	100
QY	811	ATCACCAAGATCGGCCGAGAACCCCTACAACACCCCGTGTTCCTCCATCAAGAAGAG	870
Db	101	IleLeuYsIleGlyProGluAsnProTyrAsnThrProValPheAlaIleYsYsYs	120
QY	871	GACAGCAACAGTGGCGGCACTGGTGGACTTCCGCGAGCTGACAAAGCGCACCCAGAC	930
Db	121	AspSerThrTystrpArgYsLeuValAsnPheArgGluLeuAsnYsArgThrGlnAsp	140
QY	931	TTCTGGAGGTGAGCTGGGCATCCGCCACCCCGCGGCTGAAGAAGAAAGAGCGTG	990
Db	141	PheTrpGluValGlnIleuGlyIleProHisProAlaGlyLeuYsYsYsSerVal	160
QY	991	ACCGTCTGGAGTGGGCAQCCTTACTTCAGCGTCCCGCTGACGAGACTTCCGCAAG	1050
Db	161	ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluAspPheArgYs	180
QY	1051	TACACCGCTTACCATCCCGCATCAACAGCAGACCCCGCATCCGCTACCAAGTAC	1110
Db	181	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGlnTy	200
QY	1111	AACGTCTGCTCGCGGCTGGAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAG	1170
Db	201	AsnValLeuProGlnGlyTrpYsGlySerProAlaIlePheGlnSerSerMetThrYs	220
QY	1171	ATCTCGAGCCCTTCGCGCCGCCAACCCCGAGATCGTATCTACAG-----GCCCCC	1224
Db	221	IleLeuGluProPheArgThrYsAsnProGluIleValIleTyGlnTyMetAspAsp	240
QY	1225	CTGTAGCTGGGCGACCATCTGGAGATCGCCAGCACCGGCCCAAGATCGAGGCTGGC	1284
Db	241	LeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrYsIleGluGluLeuArg	260
QY	1285	AAGCACTGTGCGTGGGCTTCCACACCCCGCAAGAAGCACCAAGAGAGCCCGCC	1344
Db	261	GluHisLeuLeuYsTrpGlyPheThrThrProAspYsYsHisGlnYsGluProPro	280
QY	1345	TTCTCTGGATGGCTACGAGCTGCAACCCGCAAGTGAACCGTGCAGCCCATCGAGCTG	1404
Db	281	PheLeuTrpMetGlyTyGluLeuHisProAspYsTrpThrValGlnProIleGlnLeu	300
QY	1405	CCCGAAGAGAGCTGGACCGTGAAGCATCCAGAGCTGTGTGGCGAGCTGAAGTGG	1464
Db	301	ProAspYsGluSerTrpThrValAsnAspIleGlnYsLeuValGlyYsLeuAsnTrp	320

QY	1465	GCCAGCCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCC	1524
Db	321	AlaSerGlnIleTyProGlyIleYsValYsGlnLeuCysYsLeuLeuArgGlyAla	340
QY	1525	AAGGCCTTACCGACATCGTGGCCCTGACCGAGGAGGCGAGCTGGCGCGAGAAC	1584
Db	341	LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn	360
QY	1585	CCGAGATCCTCGCGAGCCCGTGCACGCGTGTACTACGACCCCAAGACCTCTGGTG	1644
Db	361	ArgGluIleLeuYsGluProValHisGlyValTyTyrAspProSerYsAspLeuIle	380
QY	1645	GCCGAGATCCAGAAGCGGCCACGACCTGAGACCTACAGATCTACAGAGCCCTTC	1704
Db	381	AlaGluIleGlnYsGlnGlyGlnTyTrpThrTyGlnIleTyGlnGluGlnTy	400
QY	1705	AGAACCTCAACAGCGGCAAGTACCAAGATGCGCACCGCCACACCAACGACGTAAG	1764
Db	401	LysAsnLeuTyThrGlyTyTyrAlaArgIleYsSerAlaHisThrAsnAspValYs	420
QY	1765	CAGCTACCGAGCGCGTGCAGAAGATCGCATGAGAGCATCGTATCTGGGGCAAGACC	1824
Db	421	GlnLeuThrGluAlaValGlnYsIleAlaGlnGluSerIleValIleTrpGlyYsThr	440
QY	1825	CCCAAGTTCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACGACTACTGG	1884
Db	441	ProYsPheArgLeuProIleGlnYsGluThrTrpGluAlaTrpTrpGluTyTrp	460
QY	1885	CAGGCCACCTGCATCCCGAGTGGGAGTTCGTGACACACCCCGCTGTGAGCTGTGG	1944
Db	461	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValYsLeuTrp	480
QY	1945	TACCGTCTGGAAGGAGCGCCATCATCGCGCGCGAGACCTTCTACGTGACGCGCGCC	2004
Db	481	TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyValaspGlyAlaAla	500
QY	2005	AACCGCGAGACCAAGATCGCAAGCGCGCTACGTGACCGAGCGCGCGCGCAAGATC	2064
Db	501	AsnArgGluThrYsYsYsGlyYsAlaGlyTyValThrAspArgGlyArgGlnYsVal	520
QY	2065	GTGAGCTTACCGAGACACCAACAGAGACCGAGCTGCAGCGCCATCCAGCTGGCCCG	2124
Db	521	ValSerLeuThrGluThrThrAsnGlnYsThrGluLeuGlnAlaIleHisLeuAlaLeu	540
QY	2125	CAGGACAGCGGAGCGAGGTGAACATCGTGACCGACAGCAGTACGCCCTTGGGCATATC	2184
Db	541	GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIleIle	560
QY	2185	CAGGCCACCGCCGACAAAGAGCGAGCGAGCTGGTGAACCAAGATCATCCAGAGCTGATC	2244
Db	561	GlnAlaGlnProAspYsSerGluSerGluIleValValAsnGlnIleGluGlnLeuIle	580
QY	2245	AGAAGGAGAAAGTGTACCTGAGCTGGTGGCGCCGCCCAAGGGCATCGCGCAACGAG	2304
Db	581	GlnYsAspYsValTyLeuSerTrpValProAlaHisGlyGlyIleGlyYsGlnGlu	600
QY	2305	CAGTCGACAAAGCTGTGTGAGCAAGGATCCCGAAGGTGCTTCTCGACGCGCATCGAT	2364
Db	601	GlnValAspYsLeuValSerSerGlyIleArgYsValLeuPheLeuAspGlyIleAsp	620
RESULT 12			
B47175			
reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1			
C:Species: human immunodeficiency virus type 1, HIV-1			
C>Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004			
R:Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.			
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993			
A:Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the			
A:Reference number: A47175; MUID:93126353; PMID:7678340			
A:Accession: B47175			
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra			
A:Molecule type: DNA			



Pred. No.:	2.28e-120	Length:	559
Score:	2751.00	Matches:	510
Percent Similarity:	95.89%	Conservative:	26
Best Local Similarity:	91.23%	Mismatches:	21
Query Match:	60.16%	Indels:	2
DB:	2	Gaps:	1
US-09-610-313B-31 (1-2463) x A47175 (1-559)			
QY	673	CCCATCAGCCCTCAGAGCCGTCGCGTGAAGCTGAAGCCCGCATGACGCGCCCAAG	732
Db	1	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	20
QY	733	GTGAAGCAGTGGCCCTCGACCGAGGAGAAGATCAAGGCCCTGACCCCATCTCCGAGGAG	792
Db	21	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu	40
QY	793	ATGGAGAAGAGGGCAAGATCACCAAGATCGGCCCGCCGAGAACCCCTACAAACCCCGCTG	852
Db	41	MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal	60
QY	853	TTCGCCATCAAGAAGAGGACACCAAGTGGCGCAAGCTGGTGACTTCGCGAGCTG	912
Db	61	PheAlaIleLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu	80
QY	913	AACAAGCGCACCCAGGACTTCTGGGAGTGCAGCTGGGATCCCCCACCAGCCCGCGCTG	972
Db	81	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	100
QY	973	AGAAGAAGAGAGCTGACCGTGTGGAGTGGAGCGCCCTACCTTCAGCTGCCCTG	1032
Db	101	LysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	120
QY	1033	GACGAGGACTTCGCGAAGTACACCGCTTCACCATCCCCAGCATCAACACGAGACCC	1092
Db	121	AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro	140
QY	1093	GGCATCCGCTACCAAGTACACGCTGCTGCCCGCGCTGGAAGGCGACCCGATCTTC	1152
Db	141	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	160
QY	1153	CAGAGCAGATGACCAAGATCTCGAGCCCTTCGCGCCCGGCAACCCCGAGATCGTGATC	1212
Db	161	GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspMetValIle	180
QY	1213	TACCAG-----GCCCCCTGTACGTGGCGAGCATCGGATCGCCAGCACCGCGCC	1266
Db	181	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr	200
QY	1267	AGATCGAGGAGTGGCGCAAGCACCTGCTCGCTGGGGCTTCAACACCCCGCACAGAAG	1326
Db	201	LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys	220
QY	1327	CACCAAGAGAGCCCTTCCTCTGTGGATGGGTACGAGCTGACCCCGACCAAGTGAC	1386
Db	221	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	240
QY	1387	GTGACGCCATCGAGTGGCCGAGAGGAGAGTGGACCGTGAACACATCCAGAGCTG	1446
Db	241	ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu	260
QY	1447	GTGGGCAAGCTGAACCTGGGCGACCATCTACCCCGCATCAAGTGGCGCAGCTGTGC	1506
Db	261	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValLysGlnLeuCys	280
QY	1507	AAGTCTCTGGCGGCCCAAGGCCCTGACGACATCGTCCCTGACCGAGAGGCCGAG	1566
Db	281	LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluAlaGlu	300
QY	1567	CTGGAGCTGGCGAGAACCGGAGATCTCTGCGGAGCCGTCACGCGCTGTACTACGAC	1626
Db	301	LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp	320

QY	1627	CCACGCAAGCAGCTGTGTGGCCGAGATCCAGAGCAGGCGCACAGCAAGTGGACCTACAG	1686
Db	321	ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGln	340
QY	1687	ATCTACCAAGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCCAAGATGGCACCGCC	1746
Db	341	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla	360
QY	1747	CACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAAGATGCCATGAGAGCATC	1806
Db	361	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle	380
QY	1807	GTGATCTGGGCAAGACCCCAAGTTCGCCCTGCCATCCAGAAGAGACCTGGGAGACC	1866
Db	381	ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla	400
QY	1867	TGGTGAACCGACTACTGGCAGGCCCACTGGATCCCGAGTGGAGTTCGTGAACACCC	1926
Db	401	TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	420
QY	1927	CCCTTGGTGAAGCTGTGTACCACTGGAGAGGAGCCCATCATCGGCCCGGAGACTTC	1986
Db	421	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe	440
QY	1987	TACGTGGAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCCAG	2046
Db	441	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	460
QY	2047	CGGGCGCGCAGAAATCGTGAGCTGACCGAGACCCAGACCAACAGAGACCGGCTGAG	2106
Db	461	LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln	480
QY	2107	GCCATCCAGCTGGCCCTGCAGGACGCGCAGAGTGAACATCGTACCACGAGCCAG	2166
Db	481	AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln	500
QY	2167	TACGCCCTGGGCATCATCCAGGCCCGCAGCCGACAAAGAGCAGAGCGAGCTGTGAAC	2226
Db	501	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln	520
QY	2227	ATCATCGACGACTGATCAAGAGGAGAGGTGTACCTGAGCTGGTGGTCCGCCACAG	2286
Db	521	IleIleGluGluLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys	540
QY	2287	GCATCGCGCGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGGCATCGCAAGGTG	2343
Db	541	GlyIleGlyGlyAsnGluValValAspLysLeuValSerAlaGlyIleArgLysVal	559
RESULT 14			
S46347			
pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)			
C:Species: simian immunodeficiency virus SIVagm			
A:Variety: isolate SAB-1			
C:Date: 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999			
R:Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.; et			
EMBO J. 13, 2935-2947, 1994			
A:Title: Mosaic genome structure of simian immunodeficiency virus from West African green			
A:Reference number: S46335; MUID:94298785; PMID:8026477			
A:Accession: S46347			
A:Status: nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-1039 <JTN>			
A:Cross-references: EMBL:U04005; NID:g466229; PIDN:AAA21505.1; PID:g466231			
A:Experimental source: isolate SAB-1; sabaeus monkey			
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993			
A:Note: this reading frame extends between two stop codons and does not begin with a start			
C:Genetics:			
A:Gene: pol			
C:Superfamily: pol polyprotein			
Alignment Scores:			
Pred. No.:	9.73e-102	Length:	1039

Score:	2350.00	Matches:	440
Percent Similarity:	74.07%	Conservative:	120
Best Local Similarity:	58.20%	Mismatches:	152
Query Match:	51.39%	Indels:	44
DB:	2	Gaps:	7
US-09-610-313B-31 (1-2463) x S46347 (1-1039)			
QY	220	TTCTTCGCGAGACCTGGCTTCCCCAGGCGC-----AAGGCCCGCGATTCCCGCAGC	273
Db	1	PhePheArg-----ValTrpProLeuGlyGlnArgGluThrGlnGluPheProSer	17
QY	274	GAGCAGAACCGCGCCAC-----AGCCCCACAGCGCGAGCTGCGAGTGGCGGC-----	324
Db	18	AspLeuHisGlnThrAsnSerProAsnGlyThrGlyLeuGlnAlaGlyGlyLys	37
QY	325	-----GACAACCCCGCAGCGAGGCC-----	345
Db	38	LeuValCysArgGlnThrSerAspGlnArgThrArgAlaArgSerSerAsnSerPro	57
QY	346	-----GGCGCCGAGCGC-----	357
Db	58	ValLysAlaValCysSerGlyGluThrAlaGluThrAlaValAlaLysProLeuAla	77
QY	358	-----CAGGCGACCTGAACTTCCCCAGATCACCTGTGGCGAGCGCCC	402
Db	78	ThrThrGluProLeuArgGlyGlyLeuGlnLeuProGlnValSerLeuTrpArgPro	97
QY	403	CTGGTGAGCATCAAGTGGCGGCCAGATCAAGGAGCCCTGTGGACACGCGCGCGAC	462
Db	98	MetCysThrValTyrIleGluGlyGlnLysValThrAlaLeuLeuAspThrGlyAlaAsp	117
QY	463	GACACCGTGTGGAGGAGATGAGCTGCCCGCGCAAGTGGAGGCCCTGATGATCGCGCGC	522
Db	118	AspSerValIleGlnGlyIleGluLeuGlyAspAsnTrpLysProArgIleIleGlyGly	137
QY	523	ATCGCGCGTTCATCAAGTGGCGCGAGTACACAGATCCTGTGATCGAGATCTCGGCGAAG	582
Db	138	IleGlyGlyCysIleAsnValLysAlaTyrHisAsnGlnGluValLysIleGluAspLys	157
QY	583	AAGGCCATCGGCGCGTGTGATCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	642
Db	158	ThrCysLysAlaThrIleLeuValGlyGluThrProValAsnIleIleGlyArgAsnVal	177
QY	643	CTGACCCAGCTGGCGTGCACCTCGAACTTCCCGCATCAGCCCGCATCGAGCGCGCGCG	702
Db	178	LeuAlaGlnLeuGlyValThrLeuAsnLeuThrGlnArgGluIleGluProIleLysVal	197
QY	703	AGCTGAAGCCCGGCGATGAGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	762
Db	198	HisLeuLysProGlyGlnAspGlyProArgIleArgGlnTrpProLeuSerLysGluLys	217
QY	763	ATCAAGCCCTGACCGCATCTGGAGGAGATGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	822
Db	218	IleGluAlaLeuLysAlaIleCysGluAspLeuGlnLysGlnGlyHisLeuGluArgIle	237
QY	823	GGCGCGGAGAACCCCTACAACACCCCGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAG	882
Db	238	GlyProGluAsnProTyrAsnThrProValPheAlaIleArgLysLysAspLysThrGln	257
QY	883	TGCGCGAAGCTGTGGACTTCCCGAGCTCAACAGCGCACCCAGGACTTCTGGGAGGTG	942
Db	258	TrpArgIleLeuMetAspPheArgGlnLeuAsnLysSerThrGlnAspPheGlnGluVal	277
QY	943	CAGCTGGGCATCCCG	1002
Db	278	GlnLeuGlyIleProHisProAlaGlyLeuGlnGlnArgGlnGlnIleThrValLeuAsp	297
QY	1003	GTGGCGGAGCGCTTACTTTCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1062
Db	298	IleGlyAspAlaTyrPheSerCysProLeuAspProAspPheGlnLysTyrThrAlaPhe	317
QY	1063	ACCATCCCGCAGCATCAACACGAGACCCCGCGCATCCGCTTACCAGTACAACGTGCTGCC	1122
Db	318	ThrIleProSerValAsnAsnArgGluProGlyIleArgTyrGlnTyrLysValLeuPro	337
QY	1123	CAGGCGTGAAGGCGAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCCTCGAGGCC	1182
Db	338	GlnGlyTrpLysGlySerProThrIlePheGlnThrThrAlaAsnLysIleLeuGlnGlu	357
QY	1183	TTCCGCGCGCGCAACCCCGAGATCGTATCTACAG-----GCCCGCTGTACGTGGCG	1236
Db	358	PheArgGlnLysAsnProAspValAspIleTyrGlnTyrMetAspAspMetLeuIleAla	377
QY	1237	AGCGACCTGAGATCGCGCAGCAGCGCCCAAGATCGAGAGCTGCCAAGACCTCTG	1296
Db	378	SerAspArgProLysAlaGluHisLeuValMetValGlnGlnLeuArgAspTyrLeuGlu	397
QY	1297	CGCTGGCGCTTCCACCCCGCAGAACAGCAGAGGAGCCCGCTTCTCTGTGATG	1356
Db	398	ThrTrpGlyPheLysThrProGluLysLysPheGlnLysAspProTyrLeuTrpMet	417
QY	1357	GGCTACGAGTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAAGAG	1416
Db	418	GlyTyrGluLeuTyrProLysLysTrpGlnLeuGlnGluIleThrLeuProGluArgGlu	437
QY	1417	AGCTGGACCGTGAACGACATCCAGAGCTGGTGGCGAGCTGAACCTGGCGCAGCATC	1476
Db	438	GluTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle	457
QY	1477	TACCCCGCGCATCAAGTGGCGCGCTGTGCAAGCTGTGGCGCGCGCGCGCGCGCGCG	1536
Db	458	TyrThrGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyAlaArgProLeuThr	477
QY	1537	GACATCGTGCCTGACCGGAGGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTG	1596
Db	478	GluIleValGlnTrpThrGluGluAlaGluLeuGluLeuGluAsnArgGlnIleLeu	497
QY	1597	CGCGAGCCGTGACCGCGTGTACTAGCACCCAGCAGCAGCTGGTGGCGAGATCCAG	1656
Db	498	ArgGlnLysGlnGlnGlyGlnTyrAspProAlaLeuProLeuArgAlaLysValLeu	517
QY	1657	AAGCAGCGCCCGCAGCAGTGGACCTACAGATCTTACCGAGCGCGCTTCAAGAACCTGAAG	1716
Db	518	LysLeuGlyAspGlyGlnTrpGlyTyrGlnIleTyrGlnProGluAsnLysIleLeuLys	537
QY	1717	ACCGCAAGTACCCCAAGATGCGCGCGCCACCAACGACGTAAGTGAAGCTGACCGAG	1776
Db	538	ValGlyLysTyrAlaLysIleLysThrAlaHisThrAsnGluLeuArgMetLeuAlaGly	557
QY	1777	GCGTGCAGAGATCGCCATGGAGAGCATCGTATCTGGGCGCAGACCCCGAGTTCCGC	1836
Db	558	LeuValGlnLysIleGlyLysGluSerIleValIleTrpGlyGlnIleProIleMetGlu	577
QY	1837	CTGCCATCCAGAGGAGACCTGGAGACCTGTGGACCGACCTACTTGGCAGCGCCCTGG	1896
Db	578	LeuProValGluArgGluLeuTrpGluTrpSerAspTyrTrpGlnValThrTrp	597
QY	1897	ATCCCGAGTGGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGCTGAGCTGGAG	1956
Db	598	IleProGluTrpGluMetValSerThrProGlnLeuIleArgLeuTrpTyrLysLeuVal	617
QY	1957	AAGAGCCCATCATCGCGCGCGAGACCTTCTAGTGACCGCGCGCGCGCGCGCGCGGACC	2016
Db	618	LysAspProIleProGlyGluAlaValTyrValAspGlyAlaAlaAsnArgAsnSer	637
QY	2017	AAGTGGCAAGCGCGCTACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2076
Db	638	LysGluGlyLysAlaGlyTyrLeuThrAspArgGlyAspGlnLysValAlaLeuGlu	657
QY	2077	GAGACCAACCAAGAGAGCGAGCTGCAGCGCCATCCAGCTGGCGCTGCGAGCAGCGCG	2136
Db	658	AsnThrThrAsnGlnLysAlaGluLeuGluAlaIleLeuLeuAlaLeuArgAspSerGly	677
QY	2137	AGCAGGTGAACATCGTGACCGCAGCAGCTAGCCCTGGCGCATCATCCAGGCCCGCC	2196



Db 678 SerLysValAsnIleIleThrAspSerGlnTyrAlaMetGlyIleIleAlaGlyGluPro 697  
QY 2197 GACNAGCGAGCGAGCTGTGTGACCAACAGATCATCGACAGCTGATCAAGAGGAGAG 2256  
Db 698 ThrGluSerAspAsnAsnIleValGlnGlnIleIleGluGluIleIleLysGluAla 717  
QY 2257 GTGTACTCTGAGTGGTGGCCGCCCAACAAAGGGCATCGCGCGCAACAGCAGATCGACAAG 2316  
Db 718 ValTyrIleAlaTrpValProIleHisLysGlyValGlyGlyAsnGluGluIleAspLys 737  
QY 2317 CTGGTGAGCAAGGCGATCCGCAAGGTGTTCTCTGGAGCGGATCGAT 2364  
Db 738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753

RESULT 15  
S53092  
pol polyprotein - human immunodeficiency virus type 2  
C:Species: human immunodeficiency virus type 2, HIV-2  
C:Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
C:Accession: S53092  
R:Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.  
submitted to the EMBL data library, March 1995  
A:Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu  
A:Reference number: S53091  
A:Accession: S53092  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1055 <BEC>  
A:Cross-references: UNIPROT:O73194; EMBL:Z48731  
C:Superfamily: pol polyprotein  
C:Keywords: polyprotein

Alignment Scores:  
Pred. No.: 3.36e-95 Length: 1055  
Score: 2209.00 Matches: 417  
Percent Similarity: 70.60% Conservative: 128  
Best Local Similarity: 54.02% Mismatches: 186  
Query Match: 48.31% Indels: 41  
DB: 2 Gaps: 10

US-09-610-313B-31 (1-2463) x S53092 (1-1055)

QY 156 GGGCTGCTGGAAGTGGCGGAAGGAGGGCCACCAGATGAAGGACTGCAC---CGAGCGCCA 212  
Db 3 GlyLeuLeuGluMetTrpGln-----AspArgThrTyrHisGlyLysValPro 18  
QY 213 GGCCTAACTT-----CTTCGG 227  
Db 19 ArgGlnThrGlyGlyPheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38  
QY 228 CGAGGACCTGGCTTCCCCCAGGGCAAGGCCCGAGTTCCTCCAGCGAGCAAA----- 281  
Db 39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerGly 58  
QY 282 CCGCGGCAACAGCCCAACAGCCGCGA-----GCTGCAAGTGGCGGGGCAACCCCGG 335  
Db 59 ProThrGlyGluIleTyrAlaAlaArgGluLysAlaGluGlyAlaGluArgGluThrIle 78  
QY 336 CAGCGAGCGCGCGC-----CGAGCGCCA-----GGGCACCTCTGAACCTT----- 374  
Db 79 GlnArgGlyAspArgGlyLeuAlaAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98  
QY 375 -----CCCCAGATCACCTGTGGCAGCGCCCGCTGGTGGAGCATCAA 416  
Db 99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValValThrAlaHi 118  
QY 417 GGTGGCGGCGCAGATCAAGAGGCTCTGTGGACACCGCGCGCGCAGCACACCTGTCTGGA 476  
Db 118 stIleGluGlyGlnProValGluValLeuLeuAspThrGlyAlaAspSerIleValAl 138  
QY 477 GAGATGAGCTCCCGCGCAAGTGGAGCCCAAGATGATCGCGCGGATCGCGGCTTCAT 536  
Db 138 aGlyIleGluLeuGlySerAsnTyrSerProLysIleValGlyGlyIleGlyPheI 158

QY 537 CAAGGTGCGCCAGTACGACACGATCTGTATCGAGATCTCGGCGCAAGAGCCCATCGGCAC 596  
Db eAsnThrLysGluTyrLysAsnValGluIleGluValLeuGlyLysArgValArgAlaTh 178  
QY 597 CGTGTGATCGCCCGACCCCGTGAACATATCGCGCGCAACATGCTGACCCAGCTGGG 656  
Db rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuG 198  
QY 657 CTGCACCTGAACTTCCCATCATCAGCCCCATCAGACCGGTGCGCGTGAAGCCCGCG 716  
Db yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProG 218  
QY 717 CATGACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAATCAAGCCCTGAC 776  
Db yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeu 238  
QY 777 CGCCATCTCGAGGAGATGGAAGAGGAGGCAAGATCACAAAGATCGGCGCCCGAGAACCC 836  
Db sGluIleCysGluLysMetGluArgGluGlyGlnLeuGluAlaProProThrAsnPr 258  
QY 837 CTACAAACCCCGTGTTCGCCCATCAAGAAGAGGAGACACCAAGTGGCGCAAGCTGT 896  
Db oTyrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysIleTrpArgMetLeu 278  
QY 897 GCACTTCCCGGAGCTGAACAAGCGCACCCAGCACTTCTGGAGAGTGCAGCTGGGCATCCC 956  
Db eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyIlePr 298  
QY 957 CCAACCCCGCGCGTGAAGAAGAAGAGCGTGCACCGTGTGGAGCGTGGCGCGCCCTA 1016  
Db oHisProAlaGlyLeuAlaLysLysArgIleThrValLeuAspValGlyAspAlaTy 318  
QY 1017 CTTACAGCTGCCCCGACGAGACTTCCGCAAGTACACCGCTTCCACCATCCCCAGCAT 1076  
Db rPheSerIleProLeuHisGluAspPheArgGlnTyrThrAlaPheThrLeuProSerVa 338  
QY 1077 CAACAAACGAGACCCCGGCATCGCTACCAAGTACAACTGTCTGCCCGCAGGGTGGAGGG 1136  
Db lAsnAsnAlaGluProGlyLysArgTyrIleTyrLysValLeuProGlnGlyTrpLysG 358  
QY 1137 CAGCCCGCATCTTCCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCCGCAA 1196  
Db ySerProAlaIlePheGlnTyrThrMetArgGlnValLeuGluProPheArgLysAlaAs 378  
QY 1197 CCCCAGATCGTGATC-----TACCAGGCCCTGTAGTGGCGAGCACCTGGAGAT 1250  
Db nGlnAspValIleIleIleGlnTyrMetAspIleIleLeuIleAlaSerAspArgThrAs 398  
QY 1251 CGGCCAGCACCGCGCAAGATCGAGAGCTGGCGCAACGACCTGCTGCTGGGGCTTCAC 1310  
Db pLeuGluHisAspArgValValLeuGlnLeuLysGluLeuAsnSerLeuGlyPheSe 418  
QY 1311 CACCCCGCAGAAAGCACACAGAGAGAGCCCGCTTCTGTGTGATGGCTACGAGCTGCA 1370  
Db rThrProAspGlyLysPheGlnLysAspProTyrArgTrpMetGlyTyrGluLeuTr 438  
QY 1371 CCCCAGACAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGAGTGGACCGTAA 1430  
Db pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysGluValTrpThrValAs 458  
QY 1431 CGACATCCAGAAGCTGTGGGCAAGCTGAATCTGGCGCAGCCAGATCTACCCCGGCATCAA 1490  
Db nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyrProGlyIleL 478  
QY 1491 GTGCGCCAGCTGTGCAAGCTGTGCGCGCGCAAGGCCCTACCGACATCTGTGCCCT 1550  
Db sThrLysHisLeuCysArgIleLeuArgGlyLysMetThrLeuThrGluIleGlnTr 498  
QY 1551 GACCGAGGAGCGGCTGAGCTGCCGAGAACCCGAGATCTCGCGGAGCCCGGTGCA 1610  
Db pThrGluLeuAlaGluAlaGluLeuGluAsnArgValIleLeuSerGlnGlnG 518



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 20:20:55 ; Search time 208.667 Seconds  
(without alignments)  
12088.668 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 4573

Sequence: 1 gtcagccaccatgcccga.....ggcgtagcccggtgaattc 2463

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09610313/runat.31052005.155136.15147/app.query.fasta.1.7893  
-DB=UniProt 03 -Qfmt=faetan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09610313 @CGN 1.1 725 @runat.31052005.155136.15147 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

UniProt 03.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4140	90.5	1427	2	Q9WF92 human immun
2	4136	90.4	1427	2	Q9WF89 human immun
3	4132	90.4	1428	2	Q9WF86 human immun
4	3990	87.3	1427	2	Q9WF62 human immun
5	3983	87.1	1429	2	Q6x4P8 human immun
6	3955	86.5	1425	2	Q6x4Q6 human immun
7	3948	86.3	1427	2	Q9WF96 human immun
8	3944	86.2	1430	2	Q6x4R4 human immun
9	3935.5	86.1	1427	2	Q9WF71 human immun
10	3931	86.0	1427	2	Q9WFA0 human immun
11	3928.5	85.9	1427	2	Q9WF65 human immun
12	3927	85.9	1437	2	Q9WF77 human immun
13	3919.5	85.7	1427	2	Q9WF74 human immun
14	3913.5	85.6	1427	2	Q8ADX5 human immun
15	3905	85.4	1425	2	Q9WF79 human immun
16	3902	85.3	1427	2	Q9WF98 human immun

#### ALIGNMENTS

RESULT 1

Q9WF92	17	3893.5	85.1	1429	2	Q9WF84	human immun
Q9WF92	18	3884	84.9	1428	2	Q6x4P0	human immun
Q9WF92	19	3874.5	84.7	1427	2	Q9WF68	human immun
Q9WF92	20	3838.5	83.9	1432	2	Q8ADZ9	human immun
Q9WF92	21	3835	83.9	1433	2	Q8ADE9	human immun
Q9WF92	22	3833	83.8	1434	2	Q8ADV1	human immun
Q9WF92	23	3831	83.8	1430	2	Q7SV36	human immun
Q9WF92	24	3827	83.7	1425	2	Q7SV20	human immun
Q9WF92	25	3821.5	83.6	1433	2	Q8ADN7	human immun
Q9WF92	26	3820.5	83.5	1433	2	Q8ADK5	human immun
Q9WF92	27	3818.5	83.5	1433	2	Q8ADS7	human immun
Q9WF92	28	3815.5	83.4	1433	2	Q8ADG5	human immun
Q9WF92	29	3812.5	83.4	1435	2	Q9DQ33	human immun
Q9WF92	30	3811.5	83.3	1433	2	Q8ADE1	human immun
Q9WF92	31	3806.5	83.2	1431	2	Q8ADR1	human immun
Q9WF92	32	3805.5	83.2	1433	2	Q8ADC5	human immun
Q9WF92	33	3804.5	83.2	1433	2	Q8AE32	human immun
Q9WF92	34	3803.5	83.2	1433	2	Q8ADM1	human immun
Q9WF92	35	3803.5	83.2	1433	2	Q8ADU3	human immun
Q9WF92	36	3802.5	83.2	1437	2	O41782	human immun
Q9WF92	37	3801	83.1	1434	2	Q8Q727	human immun
Q9WF92	38	3801	83.1	1435	2	Q9WLJ3	human immun
Q9WF92	39	3800.5	83.1	1433	2	Q8AD85	human immun
Q9WF92	40	3798.5	83.1	1433	2	Q7SPT9	human immun
Q9WF92	41	3794.5	83.0	1433	2	Q8ADW7	human immun
Q9WF92	42	3794.5	83.0	1436	2	Q8AD93	human immun
Q9WF92	43	3792.5	82.9	1433	2	Q8AE23	human immun
Q9WF92	44	3791.5	82.9	1433	2	Q8ADR9	human immun
Q9WF92	45	3789.5	82.9	1433	2	Q8ADH3	human immun

Q9WF92 PRELIMINARY; PRT; 1427 AA.

AC Q9WF92; 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Gag-pol polyprotein.

OS Human immunodeficiency virus 1.

OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99214383; PubMed=10196340;

RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,

RA Foley B.T., Ndung'u T.P., Rahma M., Makhema M.J., Marlink R.,

RA Essex M.,

RT "Molecular cloning and phylogenetic analysis of human immunodeficiency

RT virus type 1 subtype C: a set of 23 full-length clones from

RT Botswana."

RL J. Virol. 73:4427-4432(1999).

CC -1- SIMILARITY: Belongs to peptidase family A2.

DR EMBL; AF110975; AAD17135.1; -

DR HSSP; P04585; 1JLA.

DR GO; GO:0019012; C:virion; IEA.

DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.

DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0008907; F:integrase activity; IEA.

DR GO; GO:0008233; F:peptidase activity; IEA.

DR GO; GO:0004523; F:ribonuclease H activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0008270; F:zinc ion binding; IEA.

DR GO; GO:0015074; P:DNA integration; IEA.

DR GO; GO:0006310; P:DNA recombination; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.

DR GO; GO:0016032; P:viral life cycle; IEA.



Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1567 CTGAGTGGCCGAGAACCGAGATCTCGCGAGCCCGTGCACGCGGTACTACGAC 1626  
Db 880 LeuGluLeuAlaGluAenArgGluIleLeuArgGluProValHisGlyValTyTyAsp 899  
QY 1627 CCAGCAGAGACCTGGTGGCGGAGATCCAGAGCAGGGCCACGACGAGTGGACCTACAG 1686  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpTyTyGln 919  
QY 1687 ATCTACAGCAGGCGCTTCAAGAACCTCAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1746  
Db 920 IleTyGlnGluProPheLysAenLeuLysThrGlyLysTyAlaLysMetArgThrAla 939  
QY 1747 CACACCAACGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCATCGAGAGCATC 1806  
Db 940 HisThrAenAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
QY 1807 GTGATCTGGGCGAGACCCCAAGTCCGCTCCCATCCAGAGGAGCCTGGAGACC 1866  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
QY 1867 TGGTGGACCCACTACTCGCAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1926  
Db 980 TrpTrpThrAspTyTrpGlnAlaThrTrpIleProGluTrpGluPheValAenThrPro 999  
QY 1927 CCCTGTGTGAAGTGTGTACAGCTGGAGAGAGCCCATCATCGCGCCGAGACCTTC 1986  
Db 1000 ProLeuValLysLeuTrpTyGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
QY 1987 TAGTGGACGGCGCGCCACCGGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2046  
Db 1020 TyValAspGlyAlaAlaAenArgGluThrLysIleGlyLysAlaGlyTyValThrAsp 1039  
QY 2047 CGGCGCGCGGAGAGATCTGAGCTGACCGAGACCAACAGAGACCGAGCTGCAG 2106  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAenGlnLysThrGluLeuGln 1059  
QY 2107 GCATCCAGCTGGCCCTCGCAGGACGAGCGGAGGTGAACATCTGTGACCGACGACG 2166  
Db 1060 AlaIleGlnLeuAlaLeuGlnAenSerGlySerGluValAenIleValThrAspSerGln 1079  
QY 2167 TAGCCCTGGGCTATCTACGCGCCAGCCGACAGAGCGAGCGAGCTGGTGAACCCAG 2226  
Db 1080 TyAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAenGln 1099  
QY 2227 ATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGGTGGCCGCCCAAG 2286  
Db 1100 IleIleGlnGlnLysIleLysGlnLysValTyLeuSerTrpValProAlaHisLys 1119  
QY 2287 GGATCGGCGGCAACGAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTG 2346  
Db 1120 GlyIleGlyAenGlnIleAspLysLeuValSerLysGlyIleAenGlyValLeu 1139  
QY 2347 TTCTGCGAGCGCATCGAT 2364  
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 2

Q9WF89 PRELIMINARY; PRT; 1427 AA.

AC Q9WF89; 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,

RA Poley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana";  
RL J. Virol. 73:4427-4432(1999).  
CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110974; AAD17126.1; -;  
DR HSP; P04585; IJLA.  
DR GO: 0019012; C:Viron; IEA.  
DR GO: 0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO: 0003677; F:DNA binding; IEA.  
DR GO: 0008907; F:integrase activity; IEA.  
DR GO: 0008233; F:peptidase activity; IEA.  
DR GO: 0004523; F:ribonuclease H activity; IEA.  
DR GO: 0003723; F:RNA binding; IEA.  
DR GO: 0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO: 0005198; F:structural molecule activity; IEA.  
DR GO: 0016740; F:transferase activity; IEA.  
DR GO: 0008270; F:zinc ion binding; IEA.  
DR GO: 0015074; P:DNA integration; IEA.  
DR GO: 0006310; P:DNA recombination; IEA.  
DR GO: 0006508; P:proteolysis and peptidolysis; IEA.  
DR GO: 0006278; P:RNA-dependent DNA replication; IEA.  
DR GO: 0016332; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR001037; Integrase C.  
DR InterPro; IPR003308; Integrase Zn N.  
DR InterPro; IPR001995; Peptidase A2.  
DR InterPro; IPR009007; Pept Aspartic.  
DR InterPro; IPR001969; Pept Asp AS.  
DR InterPro; IPR010999; Retroviral matrix.  
DR InterPro; IPR000071; Retroviral p17.  
DR InterPro; IPR008916; Retroviral capsid C.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RTase.  
DR InterPro; IPR010659; RVT connect.  
DR InterPro; IPR010661; RVT thumb.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF02022; Integrase\_Zn; 1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR Pfam; PF06815; RVT\_connect; 1.  
DR Pfam; PF06817; RVT\_thumb; 1.  
DR Pfam; PF00098; zf\_CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIV1MATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS01175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS0158; ZF\_CCHC; 2.  
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
KW Polypeptide; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
KW Zinc-finger.  
SQ SEQUENCE 1427 AA; 161293 MW; 270F25DF67615C06 CRC64;

Alignment Scores:

Pred. No.: 5,85e-156 Length: 1427  
Score: 4136.00 Matches: 780  
Percent Similarity: 99.36% Conservative: 1  
Best Local Similarity: 99.24% Mismatches: 3  
Query Match: 90.44% Indels: 2  
DB: 2 Gaps: 1

US-09-610-313b-31 (1-2463) x Q9WF89 (1-1427)

QY 13 ATGCCGAGCCATGAGCCAGCCAGCCACCATCTCTGATGACGCGAGCACTTC 72



...  
360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleuMetGlnArgSerAsnPhe 379  
73 AAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGCAAGAGAGGGCCACATCGCCCGC 132  
380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
133 AACTGCGCGCCCCCGCAGAAAGGGCTCTGGAAGTGGCGCAAGAGGGCCACCAAGATG 192  
400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
193 AAGGACTGCACCGAGCGCCAGGCGCAACTTCTCCGCGAGGACTGCGCTTCCCGCCAGGGC 252  
420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439  
253 AAGGCCGCGAGTTCCCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACCGCCGAGCTG 312  
440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459  
313 CAGGTGCGCGCGCAGAACCCCGCAGCGAGGCGCGCGCCGAGCGCCAGGGCACCTGAAC 372  
460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479  
373 TTCCCCCAGATCACCTGTGGCAGCGCCCTCTGTGAGCATCAAGTGGCGCGCCAGATC 432  
480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnIle 499  
433 AAGGAGCCCTGCTGGACACCGCGCCGACACACCGTCTGGAGAGATGAGCTGCC 492  
500 LysGluAlaLeuLeuAspThrGlyAlaAspAepThrValLeuGluGluMetSerLeuPro 519  
493 GGCAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCAAGTGGCGCCAGTAC 552  
520 GlyLysTrpLysProLysMetIleGlyIleGlyLysLysAlaIleGlyThrValLeuIleGlyPro 539  
553 GACCAGATCTGTATCAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCC 612  
540 AspGlnIleuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559  
613 ACCCCGCTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGCTGACCCCTGAATTC 672  
560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579  
673 CCCATCAGCCCATTCAGACCGTCCCGCTGAAGCTGAAGCCCGGCGATGACCGCCCAAG 732  
580 ProfileSerProfileGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
733 GTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792  
600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619  
793 ATGGAGAAGGGGCAAGATCAACAAGATCGGCCCGCAGAACCCCTPACAACACCCCGTG 852  
620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTrpAsnThrProVal 639  
853 TTCCGCCATCAAGAAGAACGACACCAAGTGGCGCAAGCTGTGTGACTTCGCGAGCTG 912  
640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
913 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCGCCCGCGCTG 972  
660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
973 AAGAAGAAGACGCTGACCGTGTGGACGTGGCGAGCGCTTCTTACGCGTGGCCCTG 1032  
680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeu 699  
1033 GACGAGACTTCGCGAAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACCC 1092  
700 AspGluAspPheArgLysTrpAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
1093 GGCATCCGCTACCAAGTACACGCTGCTGCCAGGCTGGAAGGGCAGCCCGACATCTTC 1152

720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
1153 CAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGGCGCCGCAACCCCGAGATCGTATC 1212  
740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaAsnProGluIleValIle 759  
1213 TACCAG-----GCCCCCTGTACGTGGCGCAGCGACTCGAGATCGCCAGCACCCGCGCC 1266  
760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
1267 AAGATCGAGGAGCTGCGCAAGCACCTGTCTGCGCTGGGGTTCCACCAACCCCGCAAGAAG 1326  
780 LysIleGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
1327 CACCAGAAGAGCCCTTCTTGTGGATGGGTAGAGCTGACCTGCACCCCGCAGATGGACC 1386  
800 HisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
1387 GTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACCGATCCAGACATCCAGAAGCTG 1446  
820 ValGlnProIleLeuLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
1447 GTGGCAAGCTGAACCTGGGCCAGCCAGATCTATCCCGGCATCAAGGTGCGCCAGCTGTGC 1506  
840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
1507 AGCTGCTCGCGCGCCCAAGGCCCTGACCGCATGCTGCTCCCTGACCGAGGAGCGCGAG 1566  
860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
1567 CTGGAGCTGCGCAGAACCGCAGATCTCTGCGGAGCCCGTGCACCGCTGACTACTACGAC 1626  
880 LeuGluLeuAlaGluAsnArgGluIleLeuArgLysGluProValHisGlyValTyrTrpAsp 899  
1627 CCCAGCAAGGACCTGTGTGCGCCGAGATCCAGAGACGAGGGCCACGACAGTGGACCTAC 1686  
900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
1687 ATCTACGAGGCTTCAAGAACCTGAAGACCGCGCAAGTACCGCCCAAGATGCGCAGCGCC 1746  
920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
1747 CACACCAACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCCATCGAGAGATC 1806  
940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
1807 GTGATCTGGGCAAGACCCCAAGTTCCCGCTGCCATCCAGAAAGAGACCTGGGAGACC 1866  
960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
1867 TGGTGACCGACTACTGGCAGGCCCATCTGGATCCCGAGTGGGAGTTCGTGAACACCC 1926  
980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
1927 CCCTGCTGAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCCGAGACCTTC 1986  
1000 ProLeuValLysLeuTrpLysGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
1987 TAGTGTGAGCGCGCCCAACCGCAGACCAAGATCGCAAGACCCCGCTACCTGACCGAC 2046  
1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
2047 CGGGCGCGCAGAGATCTGTAGCTTGACCGAGACCCACCAACAGAGACCGAGCTGCAG 2106  
1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
2107 GCATCCAGCTGCGCCCTGCAGGACAGCGCAGCGAGGTGAACATCGTCGCCGACCGAC 2166  
1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
2167 TAGCCCTGGGCATCATCCAGCGCCCGCAGCAAGAGGAGAGCGAGCTGTGTGAACAG 2226  
1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099



Db 580 ProileSerProileGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
QY 733 GTCAACAGTGGCCCTGACCGAGGAGAGATCAAGCCCTGACCCCTGCGAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysLeuLysAlaLeuThrAlaLeuLysGluGlu 619  
QY 793 ATGCAAGAGGAGGCGCAAGATCACCAAGATCGGCCCGAGAACCCCTACAAACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysLeuThrLysLeuGlyProGluLysProLysThrProVal 639  
QY 853 TTCGCCATCAAGAAGAGACACCAAGTGGCGCAAGTGGTGAAGTTCGCCGAGCTG 912  
Db 640 PheAlaLeuLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
QY 913 ACAACGGCACCCAGACATCTGGGAGGTGCAGCTGGGCATCCCCACCCCGCGCGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyLeuProAlaGlyLeu 679  
QY 973 AAGAAGAAGAGCGCTGACCGTCTGGACGTGGCGCGCCCTACTTCAGCGTGGCCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaLysPheSerValProLeu 699  
QY 1033 GACGAGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACCCCC 1092  
Db 700 AspGluAspPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
QY 1093 GGCATCCGCTACCAAGTACAGCTGTCGCCCGCGCGCAACCCCGAGATCGTGATC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
QY 1153 CAGACGAGATGACCAAGATCTCTGGAGCCCTTCGCGCGCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysLysLysLeuGluProPheArgAlaArgAsnProGluIleValIle 759  
QY 1213 TACCAG-----GCCCCCTGTGTAGTGGGAGCGACTGAGATCGGCAGACCCGCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
QY 1267 AGATCGAGAGTGGCAGACCTGCTGCGTGGGGCTTACACCCCGCGAGAG 1326  
Db 780 LysIleGluLeuLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
QY 1327 CACCAAGAGGAGCCCTCTCTGTGGTGGGTAGCGAGCTGACCCCGCAAGTGGACC 1386  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
QY 1387 GTGACGCCATCGAGTGGCCGAGAGAGAGAGTGGACCGTGAACACATCCAGAGCTG 1446  
Db 820 ValGlnProileGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
QY 1447 GTGGGCAAGCTGAACCTGGGCGCAGATCTACCCGGCATCAAGTGGCGAGCTGTC 1506  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuLys 859  
QY 1507 AAGCTGTGCGCGCGCCAGGCGCTGACCGACATCGTGGCCCTGACCGAGAGGCGCGAG 1566  
Db 860 LysLeuLeuArgLysAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1567 CTGAGCTGGCCGAGAACCGAGATCTCTGCGAGCCCGTGCACGGCGTGTACTACGAC 1626  
Db 880 LeuGluLeuAlaGluAspArgGluIleLeuLeuArgGluProValHisGlyValTyrTrpAsp 899  
QY 1627 CCCAGCAGGACTGTGGCCGAGATCCAGACGAGGCGCACACAGTGGACCTACCGAG 1686  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
QY 1687 ATCTACAGAGGCGCTTCAAGAACCTCAAGACCGCAAGTACGCCAAGATCGCCACCGCC 1746  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
QY 1747 CACACCAACGAGTGAAGCAGCTGACCGAGGCGCGTGCAGAAAGATCGCATGGAGAGATC 1806  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

QY 1807 GTGATCTGGGCAAGACCCCAAGTTCCCGCTGCCCATCCAGAAGGAGACCTGGGAGACC 1866  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
QY 1867 TGGTGGACCGACTACTGGCAGCCACTCGATCCCGAGTGGAGTTCGTGAACACCCCC 1926  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
QY 1927 CCCCTGGTGAAGCTGTGTACAGCTGGGAGAGGAGCCCATCATCGCCCGGAGACTTC 1986  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
QY 1987 TACGTGGACCGCGCCCAACCGCAGACCAAGATCGCAAGCGCCGCTACGTGACCCGAC 2046  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
QY 2047 CCGGGCGCGCAGAAGATCTGTAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2106  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
QY 2107 GCATCCAGCTGGCCCTGAGGACGCGCAGCGAGGTGAACATCGTGACCGACACGAG 2166  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
QY 2167 TACGCCCTGGCATCATCCAGGCCCGCAGCGCAGCGAGGTGAACATCGTGGAACCCAG 2226  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluLeuValAsnGln 1099  
QY 2227 ATCATCGACGACTGATCAAGAGAGAGGTGTACTCTGAGCTGGTGGTGGCCGCCAAG 2286  
Db 1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2287 GGCATCGCGCGCAACGAGCAGATCGACAAGCTGCTGAGCAAGGCGCATCCGAGAGTCTG 2346  
Db 1120 GlyIleGlyLysAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139  
QY 2347 TTCCTGGACGGCATCGAT 2364  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 4  
Q9WF62 PRELIMINARY; PRT; 1427 AA.  
AC Q9WF62; MEDLINE=99214383; PubMed=10196340;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.,  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RL J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110961; AAD17021.1; -.  
DR HSSP; Q70622; IHVN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.



QY 1387 GTCCAGCCCATCGAGCTGCCGAGAGGAGGAGTGGACCGTGTAACGACATCCAGAGCTG 1446  
DB 820 ValGlnProIleAsnLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
QY 1447 GTGGGCAAGCTGAACCTGGGCGACCCAGATACCCGGGATCAAGGTGGCCAGCTGTGC 1506  
DB 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuLys 859  
QY 1507 AACCTGTGGCGGCGCCAGGCGCTGACGACATCGTGCCCTCGACCGAGAGGCGCGAG 1566  
DB 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu 879  
QY 1567 CTGGAGCTGGCCGAGAACCCGAGATCTCTGGCGAGCCCGTGACCGCGCTGTACTACGAC 1626  
DB 880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp 899  
QY 1627 CCAGAGAACCTGTGTGGCCGAGATCCAGAACGAGGCGCACACAGTGGACCTACCAG 1686  
DB 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
QY 1687 ATCTACGAGAGCCCTCAAGACCTCAAGACCGGACCGGATGACCGAAGATCGCACCGCC 1746  
DB 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
QY 1747 CACCAACAGCAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATC 1806  
DB 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle 959  
QY 1807 GTGATCTGGGGCAGACCCCAAGTTCCGCTGCCATCCAGAGAGACCTGGGAGACC 1866  
DB 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
QY 1867 TGGTGACCGACTACTGGCAGGCGCCCTCGATCCCGAGTGGGAGTTCGTGACACCCGCC 1926  
DB 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluThrGluPheValAsnThrPro 999  
QY 1927 CCCTCGTGAAGCTGTGGTACCAAGCTGGAGAGGAGCCCATCATCGCGCGCCAGACCTTC 1986  
DB 1000 ProLeuValLysLeuTrpTyrGlnLeuGluGluProIleValGlyAlaGluThrPhe 1019  
QY 1987 TAGCTGGAGCGCGCGCCCAACCGCAGACCAAGATCGGAGAGCCCGGTACGTGACCGAC 2046  
DB 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
QY 2047 CGGGCGCGCAGAGATCGTGACCTGACCGACCGACACCAACAGAGACCGAGCTGCAG 2106  
DB 1040 ArgGlyArgGlnArgIleValSerLeuThrGluThrAsnGlnLysThrGluLeuGln 1059  
QY 2107 GCCATCCAGCTGGCCCTGCAGGACGCGGAGGAGTGAACATCGTGACCGACAGCCAG 2166  
DB 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln 1079  
QY 2167 TAGCCCTGGGCATATCCAGGCCCGCCAGACCGCAGAGAGCGAGCTGGTGAAACGAG 2226  
DB 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
QY 2227 ATCATCGAGAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGGCGCCGACAG 2286  
DB 1100 IleIleGlnGlnLeuIleLysLysGluValArgValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2287 GGCATCGCGCGCAACAGGAGATCCAGACGCTGGTGAAGGCGCATCCGCAAGGTGGCTG 2346  
DB 1120 GlyIleGlyLysGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
QY 2347 TTCCTGGACGGCATCGAT 2364  
DB 1140 PheLeuAspGlyIleAsp 1145  
RESULT 5  
Q6X4P8  
ID Q6X4P8 PRELIMINARY; PRT; 1429 AA.  
AC Q6X4P8;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Gag-pol fusion polypeptide (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14709249; DOI=10.1089/089922203771881220;  
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,  
RA McCutchan F.E.;  
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
in full genome analysis, is not sustained in subgenomic regions.";  
RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Harris M.E., Birx D.L.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX Kim B., McCutchan F.E.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RX Maayan S., Zeira M.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC 1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AY255825; AAP76538.1; -;  
DR HSSP; P03367; IABG.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0015074; P:DNA integration; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR001037; Integrase\_C.  
DR InterPro; IPR003308; Integrase\_Zn\_N.  
DR InterPro; IPR001995; Peptidase\_A2.  
DR InterPro; IPR009007; Peptidase\_A2.  
DR InterPro; IPR001969; Peptidase\_A2.  
DR InterPro; IPR010999; Retroviral\_matrix.  
DR InterPro; IPR000071; Retroviral\_p17.  
DR InterPro; IPR008916; Retroviral\_p17.  
DR InterPro; IPR008919; Retroviral\_capsid\_N.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; RNaseH.  
DR InterPro; IPR004777; RTase.  
DR InterPro; IPR010659; RT connect.  
DR InterPro; IPR010661; RT thumb.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF02022; Integrase\_Zn; 1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT; 1; 1.  
DR Pfam; PF06815; RVT\_connect; 1.  
DR Pfam; PF06817; RVT\_thumb; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.

DR PRINTS; PR00234; HIV1MATRIX.  
 DR SMART; SM00343; ZNF C2HC; 2.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS50175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS50158; ZF C2HC; 2.  
 DR AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferrase; Zinc;  
 KW Zinc-finger.  
 FT NON\_TER 1  
 SO SEQUENCE 1429 AA; 161586 MW; D0B9CD124DC64E19 CRC64;

Alignment Scores:  
 Pred. No.: 6.56e-150 Length: 1429  
 Score: 3983.00 Matches: 743  
 Percent Similarity: 97.47% Conservative: 27  
 Best Local Similarity: 94.05% Mismatches: 14  
 Query Match: 87.10% Indels: 6  
 DB: 2 Gaps: 2

US-09-610-313B-31 (1-2463) x Q6X4P8 (1-1429)

Qy	13	ATGCCGAGGCGCATGAGCCAGCCAGCCAGCCCAACATCTCTGATCGAGCGAGCAACTTC	72
Db		LeuAlaGluAlaMetSerGlnAlaAsnAlaAsnIleMetMetGlnLysSerAsnPhe	377
Qy	73	AAGGGCCCAAGCCATCATCAAGTGTTCAACTGCGGCAAGAGGGCCACATCGCCGC	132
Db		LysGlyProLysArgIleLeuLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg	397
Qy	133	AACCTGCGCGCCCGCCCAAGAGGGCTGCTGGAGTGCGCGCAAGAGGGCCACCATG	192
Db		AsnCysArgAlaProArgLysArgGlyCysTrpLysCysGlyLysGluGlyHisLeuMet	417
Qy	193	AAGGACTGCACCGAGCGCCAGCCCAACTTCTTCGCGAGGACCTCGGCTTCCCGCAGGC	252
Db		LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGly	437
Qy	253	AAGGCCCGAGTTCCTCCAGGAGCAGAAACCGGCCCAAGCCCAACAGCGCGAGCTG	312
Db		GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu	457
Qy	313	CAGTGGCGCGCAACACCCCGCAGCGAGCGCGCGCCCGCCAGCGCCAGGC	363
Db		GlnValArgGlyAspAsnProLysSerGluAlaGlyAspGluArgGlnGlyAlaLeuGln	477
Qy	364	---ACCTGAACTTCCCGCCAGATCACCTCTGGCAGCGCCCTCTGTGAGCATCAAGTG	420
Db		ArgThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle	497
Qy	421	GGCGCCAGATCAAGGAGCGCTGCTGGACACCGCGCCCGCCAGCACACCGTCTGGAGG	480
Db		GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu	517
Qy	481	ATGAGCTCGCGCAAGTGGAGCCCAAGATGATCGCGGCATCGCGCGCTTCATCAAG	540
Db		IleAsnLeuProGlyLysTrpLysProLysMetIleGlyLysIleGlyPheIleLys	537
Qy	541	GTGCGCCAGTACGACCATCTGATCGAGATCTCGCGCAAGAGGCATCGGCACCGTG	600
Db		ValArgGlnTrpAspGlnIleLeuIleGluLeuLysCysGlyLysAlaIleGlyThrVal	557
Qy	601	CTGATCGGCCCGCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC	660
Db		LeuValGlyProThrProValAsnIleGlyArgAsnMetLeuThrGlnLeuLysCys	577
Qy	661	ACCTGAACTTCCCATCGCCCATCGACACCGTGCCTGAGCTGAAGCCCGCGCATG	720
Db		ThrLeuAsnPheProLysSerProIleGluThrValProValLysLeuLysProGlyMet	597
Qy	721	GACGGCCCAAGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780
Db		AspGlyProLysValLysGlnTrpProLeuThrLysGluLysIleGluAlaLeuThrAla	617

Qy	781	ATCTGCGAGGAGATGGAGAGGCGCAAGATCACCAAGATCGGCCCGAGAACCCCTAC	840
Db		IleCysGluGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyr	637
Qy	841	AACACCCCGTGTTCGCATCAAGAGAGACAGACCAAGTGGCGCAAGCTGTGGAC	900
Db		AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp	657
Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGATCCCCAC	960
Db		PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyLeuPheHis	677
Qy	961	CCCGCCGCTGAAGAAGAAGAGCGTGACCTGTGTGGAGCTGGCGAGCGCTACTTC	1020
Db		ProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTrpPhe	697
Qy	1021	AGCGTGCCTGGACGAGGACTTCCGCAAGTACACCCCTTACCANTCCCGAGCATCAAC	1080
Db		SerValProLeuAspGluGlyPheArgLysTrpThrAlaPheThrIleProSerIleAsn	717
Qy	1081	AACGAGACCCCGCATCCGCTACAGTACAACTGCTGCCCGCGGCTGGAAGGGCAGC	1140
Db		AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer	737
Qy	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCCAACCCC	1200
Db		ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro	757
Qy	1201	GAGATCGTGTATCTACCAAG-----GCCCGCTGTACCTGGCGAGGAGCTGGAGTCGC	1254
Db		GluLeuValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGly	777
Qy	1255	CAGCACCGCCCAAGATCGAGGAGCTCGCAAGCACCTGTGCTGCGCTGGGCTTCCACACC	1314
Db		GlnHisArgAlaLysIleGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThr	797
Qy	1315	CCGCAAGAAGACCCAGAGGAGCCCTTCTGTGTGGATGGGTACGAGCTGCACCCC	1374
Db		ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro	817
Qy	1375	GACAAGTGGACCTGCGAGCCCATCGAGCTCGCGAGGAGGAGAGCTGGACGTCGAACGAC	1434
Db		AspLysTrpThrValGlnProIleGlnLeuProGluLysGluSerTrpThrValAsnAsp	837
Qy	1435	ATCCAGAAGCTGTGGCAAGCTGAATGGCGCCAGCCAGATCTACCCCGGCATCAAGTG	1494
Db		IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal	857
Qy	1495	CGCGAGCTGTGCAAGCTGTGCGCGCGCCAAAGCCCTGACCGACATCTGTGCCCTGACC	1554
Db		ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThr	877
Qy	1555	GAGGAGCCGAGCTGGAGCTGGCGGAGAACCGGAGATCTCTGCGGAGCCCGTCACGGC	1614
Db		GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGly	897
Qy	1615	GTGTACTACACCCCGAGAGACCTGTGCGCGCAGATCCAGAGATCCAGAGAGGCGACGAC	1674
Db		ValTyrTrpAspProSerLysAspLeuIleAlaGluLeuGlnLysGlnGlyHisAspGln	917
Qy	1675	TGACCTTACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG	1734
Db		TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTrpAlaLys	937
Qy	1735	ATCGGACCGCCACACCAACGAGTGAACAGCTGACCGAGCCGCTGCGAGAGATCGCC	1794
Db		MetArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla	957
Qy	1795	ATGGAGGATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCTCATCCAGAGGAG	1854
Db		MecGluGlyIleValIleTrpGlyLysValProLysPheArgLeuProIleGlnLysGlu	977
Qy	1855	ACCTGGGAGACCTGTGTGGACCGGAGCTACTTGGCAGGCCACCTGGATCCCGAGTGGAGTTC	1914





418	Db	LyaspCysThrGluArgGlnAlaAsnPhePheArgGluIleLeuAlaPheProGlnGly	437
253	Qy	AAGCCCGCGAGTTCCTCCAGCAGCAGACCAACCGCCCAACGCCCCACACCGCCGAGCTG	312
438	Db	GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrAsnGlyGluLeu	457
313	Qy	CAGGTGCGCGCGCAACACCCCGCAGCAGAGCGCGCGCGCAGCGCCAGGCGCACCTGAAC	372
458	Db	GlnValArgArgAsnAsnProArgSerLysThrGlyValGluArgGlnGlyThrLeuAsn	477
373	Qy	TTCCCCCAGATACCTGTGTGCACGCCCTCTGTGAGCATCAAGGTGGCGGCACATC	432
478	Db	PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle	497
433	Qy	AAGGAGGCCCTGTCTGGACACCGCGCGCGCAGCACACCGTGTCTGGAGGAGATGAGCTGCC	492
498	Db	ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro	517
493	Qy	GGCAAGTGGAAAGCCCAAGATGATCGGGCGCATCGCGGCTTCATCAAGGTGCGCCATC	552
518	Db	GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr	537
553	Qy	GACCAGATCCTGATCGAGATCTGGCGGCAATCGGAGAGGCCATCGGCACCGTGTGATCGGCGCC	612
538	Db	AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro	557
613	Qy	ACCCCGCTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGCGCTGCACCTCAACTTC	672
558	Db	ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe	577
673	Qy	CCCATCAGCCCCATCGAGACCGTCCCGCTGAAGCTGAAGCCCGCATGAGCGGCCCAAG	732
578	Db	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	597
733	Qy	GTGAAGCAGTGGCCCCCTGACCGAGAGAGATCAAGGCCCTGACCGGCATCTGCGAGGAG	792
598	Db	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu	617
793	Qy	ATGAGAGAAGGAGGCAAGATCACCAAGATCGGCCCGCAGAACCCCTACAACACCCCGCTG	852
618	Db	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal	637
853	Qy	TTCCGCTCATAGAAGAAGGACAGCACCAAGTGGCGCAGCTGTGTGACTTCCGCGAGCTG	912
638	Db	PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	657
913	Qy	AACAAGCGCACCCAGGACTTCTGGAGGCTGACGTGGGCATTCGCCCAACCCCGCGGCTG	972
658	Db	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	677
973	Qy	AAGAAGAAGAGAGCGGTGACCGTCTGGAGCTGGCGACGCGCTACTTTCAGCGTCCCGCTG	1032
678	Db	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	697
1033	Qy	GACCAGAGACTTCCGCAAGTACACCGCTTCCACATCCCGCATCAACAACAGACACCCC	1092
698	Db	AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro	717
1093	Qy	GGCATCCGCTACCAAGTCAACAGTCTGCTGCCCGAGGGCTGGAAAGGCGACGCCAGATCTTC	1152
718	Db	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	737
1153	Qy	CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCACACCCGAGATCGTGATC	1212
738	Db	GlnSerSerMetThrArgIleLeuGluProPheArgAlaGlnAsnProGluIleValIle	757
1213	Qy	TACCAG-----GCCCGCTGTAGCTGGGACGACCTGGAGATCGCGCAGCACCGCGCC	1266
758	Db	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	777
1267	Qy	AAGATCGAGGAGTGCAGACACCTCTCGCTGGGCGCTTCACCAACCCCGCAGCAAGAG	1326

Db	778	Lys	Ile	Glu	Ileu	Leu	Arg	Gly	His	Ileu	Leu	Arg	Tyr	Glu	Leu	Arg	Tyr	Gly	Phe	Thr	Thr	Pro	Asp	Lys	Leu	797
Qy	1327	CAC	CAG	AAG	GAG	CCCC	CTT	CCT	TGT	TGA	TGG	GCT	CAG	GCT	GAC	GC	CAC	CCG	CAC	AAG	TGG	AAC	C	1386		
Db	798	His	Gln	Lys	Glu	Pro	Pro	Phe	Leu	Tyr	Met	Gly	Tyr	Glu	Leu	His	Pro	Asp	Lys	Tyr	Thr	817				
Qy	1387	GTG	CAG	CCCAT	CAG	ACT	GCG	CAG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	1446	
Db	818	Val	Gln	Pro	Ile	Gln	Leu	Pro	Glu	Lys	Asp	Ser	Tyr	Val	Asn	Asp	Ile	Gln	Lys	Leu	837					
Qy	1447	GTG	GGC	AAAG	CTG	AACT	TGG	CGC	CAG	CAG	ATC	TACC	CCG	GAT	CAAG	TGT	GCG	CAG	TGT	GCG	CAG	TGT	GCG	1506		
Db	838	Val	Gly	Lys	Leu	Asn	Thr	Pro	Ala	Ser	Gln	Ile	Tyr	Pro	Gly	Lys	Val	Arg	Gln	Leu	Cys	857				
Qy	1507	AAG	CTG	TG	CGC	CGC	CGC	CAAG	CGC	CTC	AC	CGA	CAT	CTG	TGC	CTC	TG	AC	CGA	GAG	GGC	CGA	1566			
Db	858	Lys	Leu	Leu	Arg	Gly	Ala	Lys	Ala	Leu	Thr	Asp	Ile	Val	Pro	Leu	Thr	Glu	Ile	Ala	Glu	877				
Qy	1567	CTG	GAG	TG	CCCG	CAG	AAC	CCG	CAG	ATC	TCT	CGC	GAG	CCG	CGT	TGC	AC	CGC	TGT	TACT	AG	C	1626			
Db	878	Leu	Glu	Leu	Ala	Glu	Asn	Arg	Glu	Ile	Leu	Lys	Glu	Pro	Val	His	Gly	Val	Tyr	Tyr	Asp	897				
Qy	1627	CCG	CAG	AAAG	CAC	CTG	TGG	CGC	GAG	ATC	CC	AGA	AC	CGG	CC	CAC	GAG	TGG	ACT	TACC	AG	1686				
Db	898	Pro	Ser	Lys	Asp	Leu	Ile	Ala	Glu	Ile	Gln	Lys	Gln	Lys	Asp	Gln	Tyr	Thr	Gln	917						
Qy	1687	ATC	TAC	CAG	GAG	CCCTT	CA	GA	AAC	CTC	TA	GA	AC	CGG	CA	AGT	TAC	CC	CA	AGT	TAC	CGC	AC	1746		
Db	918	Ile	Tyr	Gln	Glu	Pro	Phe	Lys	Asn	Leu	Lys	Thr	Gly	Lys	Tyr	Ala	Lys	Met	Arg	Thr	Ala	937				
Qy	1747	CAC	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	1806				
Db	938	His	Thr	Asn	Asp	Val	Lys	Gln	Leu	Val	Gly	Ala	Val	Gln	Lys	Ile	Ala	Met	Glu	Gly	Ile	957				
Qy	1807	TGT	GAT	CTG	GGC	CAAG	ACCC	CA	AGT	TCC	GCT	GC	CCAT	CC	AGA	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	1866		
Db	958	Val	Ile	Tyr	Gly	Lys	Ile	Pro	Lys	Phe	Arg	Leu	Pro	Ile	Gln	Lys	Glu	Tyr	Thr	Glu	Thr	977				
Qy	1867	TGT	TGA	ACC	CACT	TAC	GCG	AGC	CACT	TG	ATC	CCCC	CG	AGT	TGG	GAG	TCT	GT	GA	CA	CCCC	1926				
Db	978	Tyr	Tyr	Thr	Asp	Tyr	Tyr	Gln	Ala	Thr	Tyr	Ile	Pro	Asp	Tyr	Glu	Phe	Val	Asn	Thr	Pro	997				
Qy	1927	CC	CT	TGT	TGA	AG	CTG	TGT	TAC	CA	GCT	TGA	AG	AGC	CC	AT	CA	TCC	GG	CCG	CAG	CTT	1986			
Db	998	Pro	Leu	Val	Lys	Leu	Tyr	Tyr	Gln	Leu	Lys	Glu	Pro	Ile	Val	Gly	Val	Glu	Thr	Phe	1017					
Qy	1987	TAG	TG	AG	CGG	CGC	CC	CAAC	CCG	CG	AG	AC	CA	GAT	CGG	CA	AGC	CGC	CT	AC	TG	AC	2046			
Db	1018	Tyr	Val	Asp	Gly	Ala	Ala	Asn	Arg	Glu	Thr	Lys	Ile	Gly	Lys	Ala	Gly	Tyr	Val	Thr	Asp	1037				
Qy	2047	CGG	GGC	CGG	CAG	AAAT	CTG	TAG	CCCT	GAC	CG	AG	CA	CC	CA	CA	CA	CA	CA							



Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAGCGCACCCAGGACTTCTGGAGAGTTCAGCTGGGCATCCCCACCCCGCGGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAGAGAGCGTGCCTGCTGGACGTGGCGACGCTACTTTCAGGCTGCCCTG 1032  
Db 680 LysLysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeu 699  
Qy 1033 GACGAGGACTTCCCAAGTACACCGCTTACCATCCCGACATCAACAACGAGACCC 1092  
Db 700 AspGluGlyPheArgLysTrpAlaPheThrIleProSerIleAsnAsnAlaThrPro 719  
Qy 1093 GGCATCCGCTACCACTACAACGCTGCTGCCAGGCTGGAAGGACGCCACGATCTTC 1152  
Db 720 GlyIleArgTrpGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
Qy 1153 CAGAGCATGACCAAGATCTCGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759  
Qy 1213 TACCAG-----GCCCCCTCTAGTGGGACGACCTGGAGATCGGCACGACCGCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1267 AAGATCAGGAGCTGCGCAAGCACCTGCTGGCTGGGGCTTCCACACCCCGCAAGAAG 1326  
Db 780 LysIleGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799  
Qy 1327 CACCAGAGGAGCCCTCTCTGTGATGGGTACGAGCTGCACCCCGCAAGTGGACC 1386  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1387 GTGCAGCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCCAAGCTG 1446  
Db 820 ValGlnProIleGlnLeuProAspSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1447 GTGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTC 1506  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
Qy 1507 AAGTGTGCGCGCGCCAGGCGCTGACGACATCGTGCCCTGACCGAGGAGCGCGAG 1566  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879  
Qy 1567 CTGGAGCTGGCGAGAACCGGAGATCTCGCGCAGCCCGTGCACGCGCTGTACTAGCAG 1626  
Db 880 LeuGluLeuAlaGluAsnAsgluIleLeuLysGluProValHisGlyValTyrTrpAsp 899  
Qy 1627 CCCAGCAGGACCTGTGGCGGAGATCCAGAGCAGGCGCCACGACCGATGACCTACGAG 1686  
Db 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1687 ATTCACGAGGACCTTTCAGAACTGAGACCGCGCAAGTACGCGCAAGATGCGCACCGCC 1746  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1747 CACACCAACCGTGAAGCAGCTGACCGGCGCTGCAGAGATCGCCATGGAGGAGCATC 1806  
Db 940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959  
Qy 1807 GTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCCATCAGAGGAGAGACTGGGAGACC 1866  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1867 TGGTGGACCGACTACTGGCAGGCCACTGGATCCCGAGTGGGAGTTCGTGAACACCC 1926  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1927 CCCCTGTGAGCTGTGTACAGCTGAGAGAGGCCCATCATCGCGCGCGAGACCTTC 1986  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy 1987 TACTGTGACGGCGCGCCCAACCCGAGACCAAGATCGCAAGCGCGCTACGTGACCCAG 2046  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2047 CGGGCGCGCAGAGATCTGTGAGCTGACCGAGACCAACCAACGAGACCGAGCTGCAG 2106  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
Qy 2107 GCCATCCAGCTGGCGCTGAGGACGAGCGGACGAGGAGTGAACATCGTACCGACGAGCCAG 2166  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079  
Qy 2167 TACGCTTGGCATCATCTCCAGGCCCGCCGACCAAGAGCGAGAGCTGCTGTAACCCAG 2226  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
Qy 2227 ATCATCCAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGTGGCGCCGACCAAG 2286  
Db 1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
Qy 2287 GGCATCGCGCGCAACGAGCAGATCCCAAGCTGTGAGCAAGGCGCATCCGCAAGTGTG 2346  
Db 1120 GlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
Qy 2347 TTCTCTGACCGCATCGAT 2364  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 8  
Q6X4R4 ID Q6X4R4 PRELIMINARY; PRT; 1430 AA.  
AC Q6X4R4;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Gsg-poi fusion polyprotein (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14709249; DOI=10.1089/089922203771891220;  
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,  
RA McCutchan F.E.;  
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
RT in full genome analysis, is not sustained in subgenomic regions.";  
RL AIDS Res. Hum. Retroviruses 19:1125-1133 (2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Harris M.E., Birx D.L.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kim B., McCutchan F.E.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Maayan S., Zeira M.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AY255823; AAF76521.1; -;  
DR HSP; P12497; I89D.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.



```
QY 1375 GACAAGTGGACCTGACCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1434
Db 819 AsplYsrTrpThrValGlnThrIleGlnLeuProGlnLysAspSerTrpThrValAsnAsp 838
QY 1435 ATCCAGAAGCTGGTGGCAGCTGAACCTGGCGCCAGCCAGATCTACCCCGGCGATCAAGGTG 1494
Db 839 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 858
QY 1495 CGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGCCCTGACCGACATCTGTCGCCCTGAC 1554
Db 859 ArgGlnLeuGlyLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValAlaLeuThr 878
QY 1555 GAGAGGCCAGCTGGAGCTGGCCGAGACCGGAGATCTCGCGAGCGCCGCGACGCG 1614
Db 879 GluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluProValHisGly 898
QY 1615 GTGTACTTACCAACCCAGACGAGCTGTGTCGCCGAGATCTCAGAGAGCGGCGACGAC 1674
Db 899 ValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln 918
QY 1675 TGGACCTTACAGATCTTACAGAGGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCCAAG 1734
Db 919 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 938
QY 1735 ATCGGACCCGCCACACCAAGAGCTGAGCAGCTGACCGAGCGCGCTGCGAGATCGCC 1794
Db 939 ArgArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 958
QY 1795 ATGAGAGAGCTGTGATCTCGGCGAGACCCCAAGTTCGCTGCCATCCAGAGGAG 1854
Db 959 MetGluSerIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGlu 978
QY 1855 ACCTGGAGAGCTGTGAGCCGACTACTGCGAGCGCCACCTGAGTCCCGAGTGGAGTTC 1914
Db 979 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 998
QY 1915 GTGAACACCCCCCTGGTGAAGCTGTGTGTAACAGTGGAGAGGAGCCATCATCGCG 1974
Db 999 ValAsnThrProLeuValLysLeuTyrTrpGlnLeuGlnLysGluProIleValGly 1018
QY 1975 GCGAGAGCTTCTACGTGGAGCGCGCGCCCAACCGGAGACCAAGATCCGCGAGCGCGC 2034
Db 1019 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1038
QY 2035 TACGTGACCCAGCCGCGCGCGCAGAGATCGTGAAGCTGACCGAGACCAACCAAG 2094
Db 1039 TyrValThrAspLysGlyArgGlnLysValValSerLeuAsnGluThrThrAsnGlnLys 1058
QY 2095 ACCGAGCTGACGCGCATCCAGCTGGCCCTGCGAGACAGCGGCGAGCGAGTGAACATCGTG 2154
Db 1059 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1078
QY 2155 ACCGAGACCGAGTACGCGCTGGGATCATCCAGGCCCGAGCCCGAGACGCGAGCGAG 2214
Db 1079 ThrAspSerGlnTyrAlaLeuGlyIleGlnAlaGlnProAspLysSerAspSerGlu 1098
QY 2215 CTGCTGAACAGATCATCGAGAGCTGATCAAGAAGGAGAGTGTACCTGAGCTGGTG 2274
Db 1099 IleValAsnGlnIleIleGlnLeuIleLysLysGluArgValTyrLeuSerTrpVal 1118
QY 2275 CCGCGCCACAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGCATC 2334
Db 1119 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerLysGlyIle 1138
QY 2335 CGCAAGTGTCTTCTCGAGCGGATCGAT 2364
Db 1139 ArgLysValLeuPheLeuAspGlyIleAsp 1148
RESULT 9
Q9WF71 PRELIMINARY; PRT; 1427 AA.
AC Q9WF71;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
```

```
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.;
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana.";
RL J. Virol. 73:4427-4432(1999).
CC -|- SIMILARITY: Belongs to peptidase family A2.
DR EMBL; AF110964; AAD17047.1; --
DR HSPSP; Q70622; IHN.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006310; P:DNA integration; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn.
DR InterPro; IPR001995; Peptidase_A2.
DR InterPro; IPR009007; Pept_Asp_AS.
DR InterPro; IPR001969; Pept_Asp_AS.
DR InterPro; IPR010999; Retrovir_matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR010659; RVT_connect.
DR InterPro; IPR010661; RVT_thumb.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF005540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT; 1.
DR Pfam; PF06817; RVT_thumb; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIV1MATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
DR PROSITE; PS00158; ZF_CCHC; 2.
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polypeptide; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1427 AA; 161326 MW; B9FC43C29B26CFA2 CRC64;
```

Alignment Scores:



Pred. No.:	4, 96e-148	Length:	1427
Score:	3935.50	Matches:	727
Percent Similarity:	96.95%	Conservative:	35
Best Local Similarity:	92.49%	Mismatches:	21
Query Match:	86.06%	Indels:	3
DB:	2	Gaps:	2
US-09-610-313B-31 (1-2463) x Q9WP71 (1-1427)			
QY	13	ATGGCGAGCCATGAGCAGGCGCCACAGCCGCAACATCTGATGAGCGCAGCACTTC	72
Db	361	LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe	379
QY	73	AAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGCGCAAGAGAGGCCACATCGCCCGC	132
Db	380	LysGlyProArgArgileValLysCysPheAsnCysGlyLysGluGlyHisleAlaArg	399
QY	133	AACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGAGGCCACCAAGATG	192
Db	400	AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet	419
QY	193	AAGGACTGCACGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCCCTCCCGCAGGC	252
Db	420	LysAspCysThrGluArgGlnAlaAsnPheArgGluAsnLeuAlaPheProGlnGly	439
QY	253	AAGGCCCGCGAGTTCCCGCAGCAGAGAAACCGCGCCCAACAGCGCCACCGCGCAGCTG	312
Db	440	GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu	459
QY	313	CAGGTGCGCGGCACAAACCCCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	372
Db	460	GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyAsnPheAsn	479
QY	373	TTCCCCCAGATCACCTGTGGCAGCGCCCGCCCTGGTGAGCATCAGGTGGCGCGCAGATC	432
Db	480	PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle	499
QY	433	AAGGAGGCCCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC	492
Db	500	ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro	519
QY	493	GGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGCGCTTCATCAAGGTGCGCCAGTAC	552
Db	520	GlyLysTrpLysProLysIleIleGlyLysIleGlyPheIleLysValArgGlnTyr	539
QY	553	GACCATCTGATTCAGATCTCGCGCAAGAGGCCATCGCACCTGTGATCGCGCCC	612
Db	540	GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro	559
QY	613	ACCCCGTGAAACATCATCGCGCAACATGCTGACCGAGTGGCGCTGCACCTGAACTTC	672
Db	560	ThrProIleAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe	579
QY	673	CCCATCAGCCCCATCAGACCGTGGCCCGTCAAGCTGAAGCCCGGCATGACGCGCCCAAG	732
Db	580	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	599
QY	733	GTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCCCATCTGCGAGGAG	792
Db	600	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu	619
QY	793	ATGGAGAAGGAGGCAAGATCAACCAAGATCGCGCGCGCGAGAACCCCTACCAACACCCCGTG	852
Db	620	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTrpAsnThrProVal	539
QY	853	TTTCGCCATCAAGAAGGACACCAAGTGGCGCAAGCTGGTGACATTCGCGCAGCTG	912
Db	640	PheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
QY	913	ACAAGCGCACCCAGACTTCTGGAGGTGCGAGCTGGGCATCCCCACCCCGCGCGCTG	972
Db	660	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	679

QY	973	AAGAAGAAGAGAGCGGTGACCTGTGTGGACGTGGGCGAGCGCTACTTCAAGCTGCCCTG	1032
Db	680	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
QY	1033	GACGAGGACTTCCGCAAGTACACCCGCTTCAACATCCCGCAGCATCACACGAGACCCC	1092
Db	700	AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro	719
QY	1093	GCATCCCGTACACAGTACAACTGTGCTGCCAGGCGTGAAGGCGAGCCCGCAGCATCTTC	1152
Db	720	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	739
QY	1153	CAGAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATC	1212
Db	740	GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle	759
QY	1213	TTACCAG-----GCCCCCTGTACGTGGCGCAGCCTGAGATCGGCAGCCCGGCC	1266
Db	760	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
QY	1267	AAGATCGAGGAGCTGCGCAAGCACCTGTGCTGGGGCTTCAACACCCCGCAGCAAGAAG	1326
Db	780	LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys	799
QY	1327	CACCAAGAGGAGCCCCCTTCTGTGTGATGGCTACGAGCTGCACCCCGCAAGTGGACC	1386
Db	800	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	819
QY	1387	GTGAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTG	1446
Db	820	ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu	839
QY	1447	GTGGGCAAGCTCAACTGGGCGCAGCAGATCTACCCCGCATCAAGTGCAGCTGTGC	1506
Db	840	IleGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgValLysHisLeuCys	859
QY	1507	AAGCTGCTCGCGCGCGCCAAAGCCCTTGACCGCATCGTCCCTGACCGAGAGGCCGAG	1566
Db	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu	879
QY	1567	CTGGAGCTGGCGAGAACCGCGAGATCTTGGCGGAGCCCGTGCACCGCGTGTACTAGAC	1626
Db	880	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	899
QY	1627	CCAGCAAGAGACTGTGTGGCCGAGATCCAGAGCAGGCGCACGACAGTGGACCTACCAG	1686
Db	900	ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln	919
QY	1687	ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGCACCGCC	1746
Db	920	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
QY	1747	CACACCAAGCAGTGAAGCAGCTGACCGAGGCGGTGCAGAAAGATCGCCATGGAGACATC	1806
Db	940	HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleAlaMetGluCysIle	959
QY	1807	GTGATCTGGGCAAGACCCCGCAAGTTCCGCTGCCCTCATCAGAAAGAGAGACCTGGGAGAC	1866
Db	960	ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr	979
QY	1867	TGGTGACCGAGCTACTGGCAGGCGCACCTGGATCCCGCAGTGGGAGTTCGTGAACACCCC	1926
Db	980	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	999
QY	1927	CCCTCGTGAAGCTGTGGTACAGCTGGAGAGAGGCCCATCATCATCGCGCGCGAGCCTTC	1986
Db	1000	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe	1019
QY	1987	TACGTGGAGCGCGCCCAACCGCAGACCAAGATCGCAAGCCCGCGCTACGTGACCGAC	2046
Db	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp	1039
QY	2047	CGGGGCGCGCAGAAGATCTGTGAGCTGTGACCGAGACCACCAACGAGAACCGAGCTGCAG	2106

```

Db 1040 LysGlyArgGlnYsIleValSerLeuAsnGluThrAenGlnLysAlaGluLeuGln 1059
Qy 2107 GCCATCCAGCTGGCCCTGCAGCAGCGCCAGCGAGGTGAACATCGTGACCGACGACG 2166
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079
Qy 2167 TACGCCCTGGCATCATCCAGGCCAGCCGACGAGCGAGCGAGCGAGCGAGCGAGCGAG 2226
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099
Qy 2227 ATCATCCAGCAGCTGATCACAAGAGAGAGTGTACCTGAGCTGGGTGCCGCCACAG 2286
Db 1100 IleIleGluGlnLeuIleLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119
Qy 2287 GGCATCCGCGCGCACGACGACGATCACAAGCTGCTGAGCAGGCGCATCCGCAAGTCTG 2346
Db 1120 GlyIleGlyAenGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139
Qy 2347 TTCTGTGACGCGATCGAT 2364
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 10
Q9WFA0
AC Q9WFA0 PRELIMINARY; PRT; 1427 AA.
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.;
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana.";
RL J. Virol. 73:4427-4432(1999).
CC -/- SIMILARITY: Belongs to peptidase family A2.
DR EMBL; AF110978; A017162.1; -.
DR PIR; S49086; S49086.
DR HSP; Q70622; IHW.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0015074; P:DNA integration; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR001037; Integrase C.
DR InterPro; IPR003308; Integrase Zn N.
DR InterPro; IPR001995; Peptidase A2.
DR InterPro; IPR009007; Pept Aspartic.
DR InterPro; IPR001969; Pept_Asp AS.
DR InterPro; IPR010999; Retroviral_matrix.
DR InterPro; IPR000071; Retroviral_p17.
DR InterPro; IPR008916; Retrov_capsid_C.

```

```

DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR010659; RVT_connect.
DR InterPro; IPR010661; RVT_thumb.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF02022; Integrase Zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; Rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT_1; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF06817; RVT_thumb; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIVMATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
DR PROSITE; PS00158; ZF_CCHC; 2.
DR AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1427 AA; 160503 MW; 24C410797550C0F8 CRC64;

Alignment Scores:
Pred. No.: 7.47e-148 Length: 1427
Score: 3931.00 Matches: 727
Percent Similarity: 96.95% Conservative: 35
Best Local Similarity: 82.49% Mismatches: 22
Query Match: 95.96% Indels: 2
DB: 2 Gaps: 1

US-09-610-313B-31 (1-2463) x 09WFA0 (1-1427)
Qy 13 ATGCCGAGCGCCATGATGACGAGCGACGAGCGACCAACATCTCTGATGACGAGCGACCACTTC 72
Db 360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379
Qy 73 AAGGGCCCCAAGCGCATCATCAAGTCTTCACTGCGGGAAGGAGGCGCCATCGCCGC 132
Db 380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399
Qy 133 AACTGCGCGCGCCCGCGCAGAGAGGCTGCTGGAAGTGGCGGAGGAGGCGCCACCATG 192
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419
Qy 193 AAGGACTGCACCGAGCGCGCCAGCGCAACTTCTCCGCGAGCACCTTCTCCCGAGCGCTTCCCGAGCGC 252
Db 420 LysAspCysThrGluArgGlnAlaAspPhePheArgGluAsnLeuAlaPheProGlnGly 439
Qy 253 AAGCGCGCGAGTTCGCCAGCGAGAGAACCGCGCCCAACAGCCCCACCGCCGCGAGCTG 312
Db 440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerGlyGluLeu 459
Qy 313 CAGTGGCGCGCGACACACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGAAC 372
Db 460 GlnValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479
Qy 373 TCCCCCGAGATCACCTGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 432
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnVal 499
Qy 433 AAGGAGCGCGCTGCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCC 492
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspThrValLeuGluGluLeuLeuLeuPro 519
Qy 493 GCGAAGTGGAGCGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTAC 552
Db 520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyPheIleLysValArgGlnTrp 539

```





Qy	1267	AAGATCGAGGAGCTGCGCAACGACCTCTCGCTGGCGCTTCACCCACCCCGCAAGAAG	1326
Db	780	LysIleGluGluLeuArgGluHisLeuLysTrpGlyLeuThrThrProTyrLysLys	799
Qy	1327	CACCAAGAAGGCCCCCTCTCTGTGATGGGCTACGAGCTGCACCCCGCAAGTGGACC	1386
Db	800	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	819
Qy	1387	GTGCAGCCCATCGAGCTGCCGGAAGAGGAGCTGGACCTGTGAACGACATCCACAGAAGCTG	1446
Db	820	ValGlnProIleGlnLeuProAspLysAspSerTrpThrValHisAspIleGlnLysLeu	839
Qy	1447	GTGGCAAGCTGAACCTGGGCGACGACATCTACCCCGGCATCAAGTGGCGCCAGCTGTGC	1506
Db	840	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuCys	859
Qy	1507	AAGCTGTGCGCGCGCAAGGCCCTGACCGACATCTGTGCCCTGTGACCGAGAGGCGCGAG	1566
Db	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu	879
Qy	1567	CTGAGAGTGGCCGAGAACCGGAGATCCTGGCGAGGCCGTGACCGCGGTGTAAGTACGAC	1626
Db	880	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	899
Qy	1627	CCACGAAAGACCTGGTGGCGGAGATCCAGAAGCAGGGCCACGACCACTAGTGGACTACCCAG	1686
Db	900	ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln	919
Qy	1687	ATCTACGAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC	1746
Db	920	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
Qy	1747	CACACCAACGAGCTGAAGCACTGACCGAGGCCGTGCAGAAGATCGCCCATGGAGAGCATC	1806
Db	940	HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleThrMetGluSerIle	959
Qy	1807	GTGATCTGGGCGAAGACCCCAAGTCTCGCTGCCATCCAGAAGGAGACCTCGGAGACC	1866
Db	960	ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr	979
Qy	1867	TGTGTGACCGCACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC	1926
Db	980	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	999
Qy	1927	CCCTGTGTGAAGCTGTGTACCACTGGAGAGAGGCCCATCATCGCGCGCGAGACCTTC	1986
Db	1000	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe	1019
Qy	1987	TACGTGACGCGCGCCGCAACCCGAGACCAAGATCGCAAGCCGAGCCGCTACGTGACCGCAC	2046
Db	1020	TyrValAspGlyAlaAlaAsnArgLuthrLysLeuGlyLysAlaGlyTyrValThrAsp	1039
Qy	2047	CGGGCCGGCAAGAAGATCGTGAGCCTGACCGAGACCAACCAAGAAAGACCGAGCTGCAG	2106
Db	1040	LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnArgAlaGluLeuGln	1059
Qy	2107	GCCATCCAGCTGGCCCTGCAGGACAGCGGCAGCGAGTGAACATCGTGAACGACGCGCAC	2166
Db	1060	AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln	1079
Qy	2167	TACGCCCTGGGCATCATCCAGGCCCGCCGACAGAGCGAGCGAGCTGGTGAACCCAG	2226
Db	1080	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln	1099
Qy	2227	ATCATCGAGCAGCTGATCAAGAAGAGAAAGGTGTACCTGAGCTGGGTGGGTGGTGAACCAAG	2286
Db	1100	IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys	1119
Qy	2287	GGCATCGCGCAACGAGCAGATCGACAAGCTGGTGTGAGCAAGGCGCATCCCGCAAGTGTCTG	2346
Db	1120	GlyIleGlyArgAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu	1139

Qy	2347	TTCTGGACGCATCAT	2364
Db	1140	PheLeuAspGlyIleAsp	1145
RESULT 12			
Q9WF77			
ID	Q9WF77	PRELIMINARY;	PRT; 1437 AA.
AC	Q9WF77;		
DT	01-NOV-1999	(TrEMBLrel. 12, Created)	
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)	
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)	
DE	Gag-pol polyprotein.		
OS	Human immunodeficiency virus 1.		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
OX	NCBI_TaxID=11676;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99214383; PubMed=10196340;		
RA	Novitsky V.A., Montano M.A., McLane M.F., Vannberg F.,		
RA	Foley B.T., Ndung'u T.P., Rahman M., Makhena M.J., Marlink R.,		
RA	Essex M.;		
RT	"Molecular cloning and phylogenetic analysis of human immunodeficiency		
RT	virus type 1 subtype C: a set of 23 full-length clones from		
RL	J. Virol. 73:4427-4432(1999).		
RL	-1- SIMILARITY: Belongs to peptidase family A2.		
CC	EMBL; AF110967; AAD17072.1; "		
DR	HSSP; P24736; 1NCP.		
DR	GO: GO:0019012; C:virion; IEA.		
DR	GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.		
DR	GO: GO:0003677; F:DNA binding; IEA.		
DR	GO: GO:0008907; F:integrase activity; IEA.		
DR	GO: GO:0008233; F:peptidase activity; IEA.		
DR	GO: GO:0004523; F:ribonuclease H activity; IEA.		
DR	GO: GO:0003723; F:RNA binding; IEA.		
DR	GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.		
DR	GO: GO:0001598; F:structural molecule activity; IEA.		
DR	GO: GO:00016740; F:transferase activity; IEA.		
DR	GO: GO:0008270; F:zinc ion binding; IEA.		
DR	GO: GO:0015074; P:DNA integration; IEA.		
DR	GO: GO:0006310; P:DNA recombination; IEA.		
DR	GO: GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	GO: GO:0006278; P:RNA-dependent DNA replication; IEA.		
DR	GO: GO:0016032; P:viral life cycle; IEA.		
DR	InterPro; IPRO00721; Gag_p24.		
DR	InterPro; IPRO01037; Integrase_C.		
DR	InterPro; IPRO003308; Integrase_Zn_N.		
DR	InterPro; IPRO01995; Peptidase_A2.		
DR	InterPro; IPRO09007; Pept Aspartic.		
DR	InterPro; IPRO01969; Pept Asp AS.		
DR	InterPro; IPRO10999; Retrovir_matrix.		
DR	InterPro; IPRO00071; Retrovir_pl1.		
DR	InterPro; IPRO08916; Retrov capsid_C.		
DR	InterPro; IPRO08919; Retrov capsid_N.		
DR	InterPro; IPRO02156; RNaseH.		
DR	InterPro; IPRO01584; Rve.		
DR	InterPro; IPRO00477; RVTse.		
DR	InterPro; IPRO10659; RVT_connect.		
DR	InterPro; IPRO10661; RVT_thumb.		
DR	InterPro; IPRO01878; Znf_CCHC.		
DR	Pfam; PF00540; Gag_p17; 1.		
DR	Pfam; PF00607; Gag_p24; 1.		
DR	Pfam; PF02022; Integrase_Zn; 1.		
DR	Pfam; PF00075; RNaseH; 1.		
DR	Pfam; PF00665; rve; 1.		
DR	Pfam; PF00077; RVP; 1.		
DR	Pfam; PF00078; RVT_1; 1.		
DR	Pfam; PF06815; RVT_connect; 1.		
DR	Pfam; PF06817; RVT_thumb; 1.		
DR	Pfam; PF00098; zf_CCHC; 2.		
DR	PRINTS; PR00939; C2HCZNFINGER.		
DR	PRINTS; PR00234; HIVMATRIX.		
DR	SMART; SM00343; ZnF_C2HC; 2.		

DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS50175; ASP\_PROT\_RETROW; 1.  
 DR PROSITE; PS50158; ZF\_CCHC; 2.  
 KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polypeptide; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
 KW Zinc-finger.  
 SQ SEQUENCE 1437 AA; 161978 MW; 65DAA2562FC317A6 CRC64;

Alignment Scores:  
 Pred. No.: 1.08e-147 Length: 1437  
 Score: 3927.00 Matches: 736  
 Percent Similarity: 96.23% Conservative: 30  
 Best Local Similarity: 92.46% Mismatches: 18  
 Query Match: 85.87% Indels: 12  
 DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x Q9WF77 (1-1437)

Qy	13	ATGCCGAGGCGATGAGCGAGGCGCCACAGCGCCCAACATCTCTGATGCGAGCGCAACTTC	72
Db	360	LeuAlaGluAlaMetSerGlnAlaAsnSerValAsnMetMetGlnLysSerAsnPhe	379
Qy	73	AAGSGCCCAAGCGCATCATCAAGTGTCTCAACTGCGCGAGGAGGCGCCATCGCCGC	132
Db	380	LysGlyProArgAsnValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLys	399
Qy	133	AACTGCGCGCGCCCGCAGAGGGCTGTGGAAGTGGCGGCAAGAGGGCCACAGATG	192
Db	400	AsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet	419
Qy	193	AAGACTGACCGAGCGCAGGCCAACTTCTTCGCGAGGACCTGGCTTCCCGCAGGC	252
Db	420	LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly	439
Qy	253	AAGSCCGCGAGTTCGCCAGCGAGCAGAACCGCGCC-----AACAGC	294
Db	440	GluAlaArgGluPheProGlnGlnIleArgAlaSerProAsnSerThrAsnSer	459
Qy	295	CCACACCGCGAGCTGCGAGTGGCGGCGACAAACCCCGCAGCGCGCGCGAG	354
Db	460	ProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGlu	479
Qy	355	-----CGCCAGGGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGGCC	402
Db	480	GlyGlnGlyThrLeuGlnGlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgPro	499
Qy	403	CTGCTGAGCATCAAGTGGCGGCGCAGATCAAGAGGCGCTCTGGACACCGCGCGAC	462
Db	500	LeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAlaAsp	519
Qy	463	GACACCTGTGGAGGAGATGAGCTGCGCGGCAAGTGGAGCCCAAGATGATCGCGCGC	522
Db	520	AsnThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly	539
Qy	523	ATCGCGGCTTCAATCAAGTGGCGGCGCAGTACGACAGATCTCTGATCGAGATCTGGCAAG	582
Db	540	IleGlyGlyPheIleLysValArgGlnTrpAspGlnIleValIleGluIleCysGlyLys	559
Qy	583	AAGSCCATCGCACCGCTGTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATG	642
Db	560	LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet	579
Qy	643	CTGACCCAGCTGGGCTGCACCTCAACTTCCCATCAGCCCATCGAGACCGTCCCGTG	702
Db	580	LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleThrValProVal	599
Qy	703	AAGCTGAAGCCCGCATGAGCGGCCCAAGGTGAAGCAGTGGCCCTTCACCGAGAGAAG	762
Db	600	LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpThrLeuThrGluLys	619
Qy	763	ATCAAGCCCTGACCGCATCTCGAGGAGATGAGAGAGGAGGCGCAAGATCACCAAGATC	822
Db	620	IleLysAlaLeuThrGluIleCysGluGluMetGluLysGluGlyLysIleThrLysIle	639

Qy	823	GGCCCCGAGAACCCCTACAACACCCCCCGTGTTCCTCCATCAAGAGAAGAGCAGCACCAAG	882
Db	640	GlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLys	659
Qy	883	TGGCGCAAGCTGGTGGACTTCCGCGAGCTGACAAAGCGCACCCAGGACTTCTGGGAGGTG	942
Db	660	TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal	679
Qy	943	CAGCTGGGCGATCCCCCACCCCGCGCTGAAGAAGAAGAGCGTGCACCGTGTGGAC	1002
Db	680	GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAsp	699
Qy	1003	GTGGCGCAGCGCTTACTTTCAGCGTGGCCCTGAGCAGGACTTCCGCAAGTACACCGCTTC	1062
Db	700	MetGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysThrAlaPhe	719
Qy	1063	ACCATCCCCAGCATCAACACGAGACCCCGGCATCCGCTACCAAGTACACGCTGCTGCC	1122
Db	720	ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro	739
Qy	1123	CAGGCTGGAAGGCGACGCCAGCATCTTCAGAGCAGCATGACCAAGATCCTCGAGGCC	1182
Db	740	GlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluPro	759
Qy	1183	TTCCGCGCCGCCAACCCCGAGATCTGATCTACAG-----GCCCGCTGTACGTGGC	1236
Db	760	PheArgLeuGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly	779
Qy	1237	AGCGACCTGAGATCGCGCGACCGCGCCAGATCGAGGAGCTGCGCAAGCCTGCTG	1296
Db	780	SerAspLeuGluIleGlyIleArgAlaGlnIleGluGluLeuArgGluHisLeuLeu	799
Qy	1297	CGCTGGGCGTTCACACCCCGCACAAGAAGCACCAGAGAGCCCGCTTCTCTGGATG	1356
Db	800	LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMet	819
Qy	1357	GGCTACGAGTGCACCCGACAAAGTGAGCCGTGAGCCCATCGAGTGCCTCGAGGAAG	1416
Db	820	GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysAsp	839
Qy	1417	AGCTGGACCTGAAACGATCCAGAGCTGGTGGGCAAGCTGGAAGTGGCGCCGACGATC	1476
Db	840	SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle	859
Qy	1477	TACCCGCGCATCAAGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACC	1536
Db	860	TyrProGlyIleLysValArgGlnLeuLysLeuLeuArgGlyAlaLysAlaLeuThr	879
Qy	1537	GACATCTGCGCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCTG	1596
Db	880	AspValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu	899
Qy	1597	CGCGAGCCGTGACCGCGCTGTACTAGACCCAGCAAGACCTGGTGGCGGAGATCAG	1656
Db	900	LysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGln	919
Qy	1657	AACGAGGCCACGACGATGACCTACCATCTACCGAGAGCCCTTCAAGAACCTGAAG	1716
Db	920	LysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys	939
Qy	1717	ACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGACCTGTAAGAGAGCTGACCGAG	1776
Db	940	ThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThrGlu	959
Qy	1777	GCGTGCAGAAAGTCCCATCGAGAGATCGTGTATCTGGGGCAAGACCCCAAGTTCGC	1836
Db	960	AlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArg	979
Qy	1837	CTGCCATCCAGAGGAGACCTGGAGACCTGGTGGACCGACTACTGCGAGCCGACCTGG	1896
Db	980	LeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAlaThrTrp	999





Db PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499  
Qy AAGAGGCGCCCTGCGAGACCGCGGCCGACACACCGTGTGGAGGATGAGCTGCC 492  
Db ArgGlnAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519  
Qy GCGAAGTGGAGGCCAAGATGATCGCGCGCATCGCGCTTCATCAAGGTGCGCGATAC 552  
Db GlyIleAspTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrp 539  
Qy GACCAAGTCTGTAGTGGAGATCTGGCGCAAGAGCCATCGCGCACCGTCTGATCGGCC 612  
Db GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559  
Qy ACCCCGCTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGCTGCACCTGAACCTC 672  
Db ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe 579  
Qy CCCATCGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCGCGCATGGAGCGGCCCAAG 732  
Db ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
Qy GTGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTCGCGAGGAG 792  
Db ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619  
Qy ATGAGAGAGGAGGCAAGATCACCAAGATCGCGCCCGAGAACCCCTACAAACACCCCGTG 852  
Db MetGluLysGluLysIleThrLysIleGlyProGluAsnProTrpAsnThrProVal 639  
Qy TTCGCCATCAAGAGAGGAGCAGCACCGAAGTGGCGCAAGCTGTGGACTTCGCGAGCTG 912  
Db PheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy AACAAAGCGCACCGAGGACTCTGGAGGTGCGAGTGGCGATCCCGCCACCGCGCGGCTG 972  
Db AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy AAGAAGAGAGAGCGTGCACCGTGTGACGTGGCGCAGCGCTACTTTCAGCGTGCCTGTG 1032  
Db LysLysLysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeu 699  
Qy GACGAGACTTCGCGAGTACACCGCTTACCATCCCGACGATCAACACGAGACCCCGC 1092  
Db AspGluSerPheArgLysThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719  
Qy GGCATCGCTACCAAGTCAAGCTGCTGCCAGGCTGGAGGCGCAGCCCGAGATCGTGATC 1152  
Db GlyIleArgTrpGlnTrpAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
Qy CAGAGCAGCATGACCAAGATCTCGAGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759  
Qy TACCAAG-----GCCCGCTGTAGTGGAGCGACCTGGAGATCGCGCAGCACCGCGCC 1266  
Db TyrGlnTrpMetAspAspLeuTrpValGlySerAspLeuGluIleGlyGlnHisAla 779  
Qy AAGATCGAGGAGCTGCGCAGCACCTGCTGGCTGGGGCTTCACCGCCCGCAGCAAGAG 1326  
Db LysIleGluGluLeuArgGluHisLeuLysTrpGlyLeuThrThrProAspLysLys 799  
Qy CACCAAGAGAGCGCCCTCTCTGTGGATGGCTACGAGCTGCACCGCCCGCAAGTGGACC 1386  
Db HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy GTGAGCCCATCGAGCTGCCGAGAGAGAGTGGACCGGTGAACGATCCAGAGCTG 1446  
Db ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy GTGGCGAGCTGAAGTGGCGCAGCAGATCTACCGCGCATCAAGTGGCGCAGCTGTC 1506  
Db ValGlyLysLeuAsnTrpAlaSerGlnIleTrpProGlyIleArgValArgHisLeuCys 859

Qy AAGCTGTGCGCGCGCCCAAGCGCCCTGACCAACATCGTGTCCCTGACCGAGAGGCCGAG 1566  
Db LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
Qy CTGAGCTGCGCGAGAACCGCGAGATCTCTGCGCGCGCCGTCACCGCGCTGTACTACGAC 1626  
Db LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTrpTyrAsp 899  
Qy CCCAAGAGGACCTGTGTGGCGCGAGATCCAGAAAGCAGGGCCACGACCTGACCTTACCAG 1686  
Db ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrGln 919  
Qy ATCTACGAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1746  
Db IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy CACACCAACGACGAGGAGGAGCTGACCGAGCGCTGCAGAGAGATCGCCATCGAGAGCATC 1806  
Db HisThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle 959  
Qy GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGAGAGCC 1866  
Db ValIleTrpGlyThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979  
Qy TGTGAGCCGACTACTGTGCGAGCGCACCTGGATCCCGAGTGGAGTTCGTGAACACCGCC 1926  
Db TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy CCCCTGTGTGAAGCTGTGTACCGAGCTGGAGAGAGGCCCATCATCGCGCGCGAGACCTTC 1986  
Db ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019  
Qy TACGTGACGCGCGCGCAACCGCGAGACCAAGATCGCAAGGCGCGCTACGTGACCGAC 2046  
Db TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 1039  
Qy CGGCGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2106  
Db LysGlyArgGlnLysIleValProLeuThrGluThrAsnGlnLysAlaGluLeuGln 1059  
Qy GCNATCCAGCTGCGCGCGCGAGCGCGAGCGCGAGCTGAACATCGTGACGAGCGCAG 2166  
Db AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy TACCGCTGCGCATCTCCAGCGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACCCAG 2226  
Db TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099  
Qy ATCATCGAGCAGCTGTATCAAGAGGAGAGGTGTACTGTAGCTGGTGCCTGCCCGCCACAG 2286  
Db IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
Qy GGCATCGCGCGCAGCAGCAGATCGACAGCTGTGTGACCAAGGCGCATCCCGCAAGTGTG 2346  
Db GlyIleGlyLysAsnGluValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
Qy TTCCTGCGACGCGCATCGAT 2364  
Db PheLeuAspGlyIleAsp 1145  
RESULT 14  
Q8ADX5 PRELIMINARY; PRT; 1427 AA.  
ID Q8ADX5  
AC Q8ADX5  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol fusion polypeptide (fragment).  
GN Name=gag-pol;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;



QY 793 ATGGAGAGGAGGCGAAGATCACCAAGATCGGCCCGGAGAACCCCTCAACACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIle 639  
QY 853 TTCCGCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGGTGCATCTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysIleTyrArgLysLeuValAspPheArgGluLeu 659  
QY 913 AACAGCCGACCCAGGACTTCTGGAGGTGAGCTGGGATCCGCCACCCCGCGCCCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
QY 973 AAGAGAGAGAGAGCGTGCCTGCTGACCTGGCGAGCGCTACTCTCAGCGTGCCTG 1032  
Db 680 LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
QY 1033 GACGAGAGCTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAAGAGACCC 1092  
Db 700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
QY 1093 GGCATCCGCTACCAAGTACAAGTGTGCTGCCAGGGCTGGAAGGGCAGCCCGAGCATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyrLysGlySerProAlaIlePhe 739  
QY 1153 CAGAGCATGATCAGCAAGATCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTATC 1212  
Db 740 GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759  
QY 1213 TACCAG-----GCCCGCTGTACTGGCGGACGACCTGGAGATCGGCGAGCCCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779  
QY 1267 AAGATCGAGGAGCTGCGCAAGACATGCTGCTGGCTGGGCTTCCACACCCCGCAAGAG 1326  
Db 780 LysIleGluGluLeuArgGluHisLeuLeuArgTyrGlyPheThrThrProAspLysLys 799  
QY 1327 CACGAGAGGAGCCCTTCTGTGATGGCTGACGATCGAGCTGCCACCCCGAGTGCACC 1386  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTyrThr 819  
QY 1387 GTGAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGATCCAGAGCTG 1446  
Db 820 ValGlnProIleGlnLeuProGluLysAspSerThrThrValAsnAspIleGlnLysLeu 839  
QY 1447 GTGGGCAAGCTGAATGGCGCAGCAGATCTACCCCGCATCAAGGTGCGCAGCTGTC 1506  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLysCys 859  
QY 1507 AAGTGTGCGCGCGGCGCAAGCCCTGACCCAGATCGTGGCCCTGACCGAGAGCGCGAG 1566  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1567 CTGAGCTGGCGGAGAACCGGAGATCTGCGCGAGCCCGTGCACCGGCTGTACTACGAC 1626  
Db 880 LeuGluLeuAlaGluAsnArgGluLysLysGluProValHisGlyValTyrTyrAsp 899  
QY 1627 CCCAGCAAGCACTGTGGCGCAGATCCAGAAAGCAGGCGCCACGACCATGAGTGCAC 1686  
Db 900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGln 919  
QY 1687 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCGCAAGTACGCGAGATCGCGACCC 1746  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939  
QY 1747 CACACCAAGCAAGTGAAGCAGCTACCGAGCGCCCTGAGAGATCGCCATCGAGAGTATC 1806  
Db 940 HisThrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959  
QY 1807 GTGATCTGGGCAAGACCCCAAGTCTCGCTGCCCTCCATCCAGAAAGAGAGACTGGGAGAC 1866  
Db 960 IleIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluAla 979  
QY 1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCC 1926

Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
QY 1927 CCCCTGTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGGCGCCGAGACCTTC 1986  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019  
QY 1987 TACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACTGACCCGAC 2046  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039  
QY 2047 CGGCGCGCGAAGATCGTGCAGCTGACCGAGACCAAGAGACCAAGAGAGCGAGTGCAG 2106  
Db 1040 ArgGlyArgGlnLysIleIleSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
QY 2107 GCCATCCAGCTGGCGCTGCAGACAGCGCGAGCGTGAACATCGTGACCGACGCCAG 2166  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValThrAspSerGln 1079  
QY 2167 TACGCGCTGGCATCATCAGCGCCGAGCCCGAGAGAGAGAGAGAGAGAGAGAGAG 2226  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099  
QY 2227 ATCATCGAGAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2286  
Db 1100 IleIleGluGlnLeuIleLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2287 GGCATCGCGCGCAACAG 2346  
Db 1120 GlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
QY 2347 TCTCTGAGCGCATCGAT 2364  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 15  
Q9WF79 PRELIMINARY; PRT; 1425 AA.  
ID Q9WF79  
AC Q9WF79;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polypeptide.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana";  
RL J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110969; AAD17087.1; -;  
DR HSSP; P24736; INCP.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0015074; P:DNA integration; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.



Db	838	ValGlyLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys	857
Qy	1507	AAGTCTGTCGCGCGCCCAAGCCCTGACCATCTGTCGCCCTGACCCAGGAGGCGCCGAG	1566
Db	858	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProProThrGluGluAlaGlu	877
Qy	1567	CTGAGCTGCGCGAGAAACCGGAGATCTCTGCGCGAGCCGTCGACGCGGTGTACTACGAC	1626
Db	878	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	897
Qy	1627	CCCAGCAGGACCTGGTGGCGAGATCCAGAGCAGGCGCCACCCAGTGCAGCTTACCAG	1686
Db	898	ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspGlnTrpThrTyrGln	917
Qy	1687	ATCTACCAAGGACCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCACCGCC	1746
Db	918	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	937
Qy	1747	CACACCAACGAGCTGAAGCTGACCGAGCCGTCGACAGATCGCCATCGAGAGCATC	1806
Db	938	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle	957
Qy	1807	GTGATCTGGGCGACACCCCAAGTTCGCTGCCCTCCATCCAGAGGACACCTGGGAGACC	1866
Db	958	ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla	977
Qy	1867	TGGTGGACCGACTACTGGCAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC	1926
Db	978	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	997
Qy	1927	CCCTGTGTAAGCTGTGTACAGCTGGAGAGGAGCCCATCTCGGCGCGGAGACCTTC	1986
Db	998	ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe	1017
Qy	1987	TACGTGAGCGCGCGCCCAACCGGAGACCAAGATCCGCAAGCGCGGCTAGTGCACCGAC	2046
Db	1018	TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp	1037
Qy	2047	CGGCGCGGAGAGAGCTGTGACCTGACCGAGACCAACCAAGAGACCGAGCTGCAG	2106
Db	1038	LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln	1057
Qy	2107	GCCATCCAGCTGGCCCTCGAGGACAGCGGCGAGGTGAACATCGTGACCGACAGCCAG	2166
Db	1058	AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln	1077
Qy	2167	TACGCCCTGGGCATCATCCAGCCCGCCGACAGAGCGAGGAGCGAGCTGTAACCCAG	2226
Db	1078	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln	1097
Qy	2227	ATCATCGAGCAGCTGATCAAGAAGAGAGAGGTGTACTGTGAGTGGTGGCCGCCACCAAG	2286
Db	1098	IleIleGluGlnLeuIleLysGlnLysValTyrLeuSerTrpValProAlaHisLys	1117
Qy	2287	GGCATCGGCGCAACGAGCAGATCGACAAGCTGGTGAAGGCGCATCCGCAAGGTGCTG	2346
Db	1118	GlyIleGlyGlyAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu	1137
Qy	2347	TTCCTGGACGCGCATCGAT	2364
Db	1138	PheLeuAspGlyIleAsp	1143

Search completed: June 2, 2005, 05:04:08  
Job time : 290.667 secs



***This Page Blank (uspio)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 04:23:54 ; Search time 153.167 Seconds  
(without alignments)  
11117.427 Million cell updates/sec

Title: US-09-610-313B-31  
Perfect score: 4573  
Sequence: 1 gtcagccaccatgcccga.....gggctagaccggtgaattc 2463

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters: 2931222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09610313/runat\_31052005\_155138\_15193/app\_query.fasta\_1.7893  
-DB=Published Applications AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALL=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09610313 @CGN 1 1 550 @runat\_31052005\_155138\_15193  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5  
-FGAPOP=6 -FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DEEXT=7

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\* Query  
Result

No.	Score	Match	Length	DB	ID	Description
1	3570	78.1	998	16	US-10-332-413-4	Sequence 4, Appli
2	3523.5	77.1	850	16	US-10-332-413-10	Sequence 10, Appl
3	3497	76.5	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3475	76.0	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3452	75.5	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3452	75.5	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3430	75.0	1003	16	US-10-325-468-35	Sequence 35, Appl
8	3428	75.0	1003	16	US-10-325-468-23	Sequence 23, Appl
9	3427	74.9	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3427	74.9	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3419	74.8	995	15	US-10-296-734-2	Sequence 2, Appli
12	3405	74.5	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3251.5	71.1	1006	15	US-10-296-734-1471	Sequence 1471, Ap
14	3183.5	69.6	1014	14	US-10-301-661A-6	Sequence 6, Appli
15	3119.5	68.2	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3119.5	68.2	1350	16	US-10-380-641-35	Sequence 35, Appl
17	3119.5	68.2	1350	17	US-10-636-730-35	Sequence 35, Appl
18	2822.5	61.7	875	10	US-09-952-060-6	Sequence 6, Appli
19	2822.5	61.7	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2822.5	61.7	875	16	US-10-380-641-6	Sequence 6, Appli
21	2822.5	61.7	875	17	US-10-636-730-6	Sequence 6, Appli
22	2817	61.6	850	10	US-09-952-060-2	Sequence 2, Appli
23	2817	61.6	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2817	61.6	850	16	US-10-380-641-2	Sequence 2, Appli
25	2817	61.6	850	17	US-10-636-730-2	Sequence 2, Appli
26	2798.5	61.2	875	10	US-09-952-060-8	Sequence 8, Appli
27	2798.5	61.2	875	15	US-10-168-217A-8	Sequence 8, Appli
28	2798.5	61.2	875	16	US-10-380-641-8	Sequence 8, Appli
29	2798.5	61.2	875	17	US-10-636-730-8	Sequence 8, Appli
30	2793	61.1	850	10	US-09-952-060-4	Sequence 4, Appli
31	2793	61.1	850	15	US-10-168-217A-4	Sequence 4, Appli
32	2793	61.1	850	16	US-10-380-641-4	Sequence 4, Appli
33	2793	61.1	850	17	US-10-636-730-4	Sequence 4, Appli
34	2789	61.0	582	9	US-09-735-487-14	Sequence 14, Appl
35	2789	61.0	582	14	US-10-342-188-14	Sequence 14, Appl
36	2785	60.9	560	15	US-10-399-920-2	Sequence 2, Appli
37	2783	60.9	561	14	US-10-059-271-87	Sequence 87, Appl
38	2777	60.7	562	14	US-10-102-622-10	Sequence 10, Appl
39	2777	60.3	560	14	US-10-205-641-1	Sequence 1, Appli
40	2748.5	60.1	581	14	US-10-102-622-12	Sequence 12, Appl
41	2736	59.8	560	11	US-09-725-652-1	Sequence 1, Appli
42	2702	59.1	546	8	US-08-808-031A-29	Sequence 29, Appl
43	2689	58.8	979	14	US-10-271-181B-118	Sequence 118, App
44	2588	56.6	851	14	US-10-059-271-88	Sequence 88, Appl
45	2571	56.2	999	15	US-10-346-000A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-10-332-413-4  
; Sequence 4, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swanstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdoool  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
; TITLE OF INVENTION: Subtype Isolates, Their Genes nd Modifications and Derivatives T  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924



```
QY 2137 AGCAGGTGAACATGCTGACCGACGACGAGTACGCCCTCGGCGATCATCTCCAGGCCAGCC 2196
Db 641 SerGluValAlaSerThrAspSerGlnTyrAlaLeuGlyIleGlnAlaGlnPro 660
QY 2197 GACAAGAGCAGACGAGCTGGTGAACACAGATCATCGACGAGCTGATCAAGAGGAGAAG 2256
Db 661 AspArgSerGluSerGluLeuValAsnGlnIleGlnGluLeuIleGlySerGluArg 680
QY 2257 GTGTACCTGAGCTGGGTGCGCGCCCAACAGGCGATCGCGCGCAACAGCAGATCGACAAG 2316
Db 681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700
QY 2317 CTGGTGAGCAAGGCGATCCGCAAGGTGCTGCTTCTGAGCGCGATCGAT 2364
Db 701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716

RESULT 2
US-10-332-413-10
; Sequence 10, Application US/10332413
; Publication No. US20040116660A1
; GENERAL INFORMATION:
; APPLICANT: Johnstone, Robert Edward
; APPLICANT: Swanson, Ronald Ivar
; APPLICANT: Morris, Lynn
; APPLICANT: Karim, Salim Abdool
; APPLICANT: Williams, Carolyn
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H
; TITLE OF INVENTION: Subtype Isolates, Their Genes and Modifications and Derivatives
; FILE REFERENCE: 45669-281993
; CURRENT APPLICATION NUMBER: US/10/332,413
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: PCT/IB01/01208
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216,995
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: ZA 2000/3437
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: ZA 2000/4924
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-10

Alignment Scores:
Pred. No.: 2.8e-163 Length: 850
Score: 3523.50 Matches: 668
Percent Similarity: 91.05% Conservative: 34
Best Local Similarity: 86.64% Mismatches: 44
Query Match: 77.05% Indels: 25
DB: 16 Gaps: 5

US-09-610-313b-31 (1-2463) x US-10-332-413-10 (1-850)
QY 97 TGCTTCAACTGCGCGAGGAGGCCACATCGCCCGCACTGCGCGCCCGCCGCAAGAAG 156
Db 84 CysAlaThrValGlyLys-----GlyAspArgCysGlyProLeuArgTyr 99
QY 157 GGCTGCTGGAG-----TGCGCAAGAGGAGGCCACCATG 192
Db 100 AlaSerTrpArgLysGlyAspValLeuGlnGlyAspValGlyArgGlnGlyPheProSer 119
QY 193 AAGGATGACACGAGCGCCAG-----GCCAACTTCTTCGCGAGGACCTG--- 237
Db 120 HisAspValValLysArgArgProValProSerLeuHisAlaCysArgSerThrLeuGlu 139
QY 238 -----GCCTTCCCGCAGGCGCGCGCGCGAGTCCCGCGAGCAG 279
Db 140 AspProArgValProSerPheProGlnGlyProAlaArgGlnPheProSerGluGln 159
```

```
QY 280 AACCGCGCCAAACAGCCCGCCACAGCGCGAGCTGCGCGCGCGCAACACCCCGCGAGC 339
Db 160 ThrArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAspAsnProArgSer 179
QY 340 GAGCGCGCGCCAGCGCCAGCGCCAGCGCCAGCTTCCCGCCAGATCACCTCTGCGCAGCGC 399
Db 180 GluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 199
QY 400 CCCTCTGTAGACATCAAGGTGGCGCGCCAGATCAAGAGGCGCCCTGTGCGCACCGCGGCC 459
Db 200 ProLeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAla 219
QY 460 GAGCAGACCGTCTGAGGAGATGAGCTTCCCGCGCAAGTGGAGGCCCAAGATGATCGGC 519
Db 220 AspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGly 239
QY 520 GGCATCGCGCGCTTCATCAAGGTGGCGCGCCAGTACGACCGACAGATCCTGTAGATCGCGGC 579
Db 240 GlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 259
QY 580 AAGAAGGCCATCGCGCCAGCGCTGCTGATCGCGCCCGCCAGCGCGCCAGATCATCGCGCGCAAC 639
Db 260 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 279
QY 640 ATGCTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCATCAGCCCCATCGAGACCGTGC 699
Db 280 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 299
QY 700 GTCAAGCTGAGCGCGCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 759
Db 300 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluVal 319
QY 760 AAGATCAAGGCCCTCGACCGCGCATCTCGAGGAGATGGAGAGGAGGCGCAAGATCACCAAG 819
Db 320 LysIleLysAlaLeuThrAlaIleCysGluMetGluLysGluGlyLysIleThrLys 339
QY 820 ATCGCGCCCGAGAACCCCTACAACACCCCGCTTCCCGCATCAAGAGAGAGGAGCAGCAC 879
Db 340 IleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysGluAspSerThr 359
QY 880 AGTGGCGCGAGCTGGTGGACTTCCCGGAGCTGAACAGCGCGCGCGCGCGCGCGCGCGAG 939
Db 360 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 379
QY 940 GTGCGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 999
Db 380 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu 399
QY 1000 GAGCTGGCGCGCGCTTACTTTCAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1059
Db 400 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAla 419
QY 1060 TTCACCATCCCGAGCATCAACAGAGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1119
Db 420 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 439
QY 1120 CCGCGCGCTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1179
Db 440 ProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSerMetThrLysIleLeuGlu 459
QY 1180 CCCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1233
Db 460 ProPheArgAlaLysAsnProGluIleValIleTyrGlnTyrMetAlaAlaLeuTyrVal 479
QY 1234 GGCAGGCGCTGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1293
Db 480 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 499
QY 1294 CTGCGCTGGGCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1353
Db 500 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeuTrp 519
```



QY 928 GACTTCTGGAGGTGACGCTGGGATCCCCCAACCCCGCGCTGAAAGAAAGAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLysLysLysSer 260  
QY 988 GTGACCGTGTGGAGCGGAGCCCTACTTTCAGCGTCCCTGGAGGAGCTTCGCG 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArg 280  
QY 1048 AAGTACACCGCTTCCACCATCCAGCATCAACACGAGACCCCGGATCCGCTACACG 1107  
Db 281 LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln 300  
QY 1108 TACAACGTGCTGCCAGCGGTGAGGGAGCCCGGAGCATCTTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTCTGGAGCCTTCCCGCGCCGCAACCCCGAGATCGTGATCTACCAG- ----GCC 1221  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTrpGlnTrpMetAsp 340  
QY 1222 CCCCTGTACGTGGGAGCGACCTGGAGATCGCGCAGCACCCGCGCAAGATCGAGGAGCTG 1281  
Db 341 AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1282 CGCAAGCACCTGTGCTGGGCTTCCACCAACCCCGCAAGAGACCAAGAGAGGCC 1341  
Db 361 ArgGlnHisLeuLeuArgTrpGlyLeuThrTrpProAspLysLysHisGlnLysGluPro 380  
QY 1342 CCCTTCTGTGGTGGGTACGAGCTGCACCCGCAAGTGAGCGTGCGACCCATCGAG 1401  
Db 381 ProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
QY 1402 CTGCCCGAGAGGAGAGCTGACCGCTGAACACATCCAGAGCTGGTGGGCAAGCTGAAC 1461  
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1462 TGGGCCACGCCAGATCTACCCCGCATCAAGTGGCCAGCTGTCAAGCTCTCGCGGC 1521  
Db 421 TrpAlaSerGlnIleTrpProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
QY 1522 GCCAAGCCCTGACCGACATCGTCCCTGACCGAGGAGCGGCGAGCTGGAGCTGCCGAG 1581  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
QY 1582 AACCCGAGATCTCGCGAGCCCGTGCACCGCTGTACTACGACCCCGCAAGAGCACTG 1641  
Db 461 AsnArgGluLeuLysGluProValHisGlyValTrpAspProSerLysAspLeu 480  
QY 1642 GTGCCCGAGATCCAGAGCGGCGCACGACGAGTGGACCTTACAGATCTACAGAGGCC 1701  
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTrpGlnIleTrpGlnGluPro 500  
QY 1702 TTCAGACCTGAGACCGCAAGTACGCCAAGATGCCACCGCCGCAACCAAGAGCTG 1761  
Db 501 PheLysAsnLeuLysThrGlyLysTrpAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
QY 1762 AAGCAGCTGACCGAGGCGGTGCAGAGATCCCATGAGAGCATCTGTGATCTGGGGCAAG 1821  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540  
QY 1822 ACCCCCAAGTTCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTAC 1881  
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTrp 560  
QY 1882 TGGCAGCGCCTCGATCCCGAGTGGAGTTCGTGAACACCCCGCTGCTGGTGAAGCTG 1941  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
QY 1942 TGGTACAGCTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGGCC 2001  
Db 581 TrpTrpGlnLeuGluLysGluProIleValGlyAlaGluThrPheTrpValAspGlyAla 600

QY 2002 GCCAACCCGAGAGACCAAGATCGCAAGCCCGCTTACGTGACCGACCGGCGCGGAGAG 2061  
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTrpValThrAsnArgGlyArgGlnLys 620  
QY 2062 ATCTGTAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121  
Db 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTrpLeuAla 640  
QY 2122 CTGCAGGACAGCGGAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCTTGGGCATC 2181  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTrpAlaLeuGlyIle 660  
QY 2182 ATCCAGGCGCCGAGCCGCAAGAGGAGCGAGCGAGCTGTGTAACCAAGATCATCGAGCAGT 2241  
Db 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680  
QY 2242 ATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGCCCGCCCAAGGGGATCGCGGCAAC 2301  
Db 681 IleLysLysGluLysValTrpLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700  
QY 2302 GAGCAGATCGCAAGCTGGTGAGCAAGGCGCATCCGCAAGGTGCTTCTTCTGGACGCGATC 2361  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
QY 2362 GAT 2364  
Db 721 Asp 721  
RESULT 4  
US-10-634-165-9  
; Sequence 9, Application US/10634165  
; Publication No. US20050095581A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Nancy T  
; APPLICANT: Gallo, Robert C.  
; APPLICANT: Wong-Staal, Flossie  
; TITLE OF INVENTION: DETECTION OF HIV-1 DNA  
; FILE REFERENCE: 223695  
; CURRENT APPLICATION NUMBER: US/10/634,165  
; PRIOR FILING DATE: 2003-08-05  
; PRIOR APPLICATION NUMBER: 08/463,028  
; PRIOR FILING DATE: 1995-06-05  
; PRIOR APPLICATION NUMBER: 08/693,866  
; PRIOR FILING DATE: 1985-01-23  
; PRIOR APPLICATION NUMBER: 06/659,339  
; PRIOR FILING DATE: 1984-10-10  
; PRIOR APPLICATION NUMBER: 06/643,306  
; PRIOR FILING DATE: 1984-08-22  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 9  
; LENGTH: 1015  
; TYPE: PRT  
; ORGANISM: T cell leukemia-lymphoma virus (HTLV)  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (1)..(1015)  
; OTHER INFORMATION: /note= "pol protein of HTLV-III"  
US-10-634-165-9  
Alignment Scores:  
Pred. No.: 6,42e-161 Length: 1015  
Score: 3475.00 Matches: 654  
Percent Similarity: 93.72% Conservative: 33  
Best Local Similarity: 89.22% Mismatches: 28  
Query Match: 75.99% Indels: 18  
DB: 17 Gaps: 4  
US-09-610-313B-31 (1-2463) x US-10-634-165-9 (1-1015)  
QY 220 TTCTTCCGAGGACCTGGCTTCCCGGAGGAGCGCGCGAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20



```
QY 268 -----CCGAGGAGCAGAACCGCGCCCAACAGCCCCACCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGGCGACAACCCCGCAGCGAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGGCACCCGTG-----AACTTCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGGTGGCGCGCCAGATCAAGGAGGCCCTCTGGACACCGCGCGCGACGACCGCTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100
QY 472 CTGAGGAGATGACCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
QY 532 TTCAATAGGTGGCGCGCATGACGACCGAGATCCTGATCGAGATCTGCGGCGAAGAGCGCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGCGCCCCCGCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCAACCTGAACTCCCATCAGCCCATCGAGACCGTGGCGCGTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
QY 712 CCCGGCATGACGCGCCCAAGGTGAAGCAGTGGCCCTGACCCAGAGAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCGCCATCTCGGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGCGCCCGAG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
QY 832 AACCCCTTACAACACCCCGGTGTTCCGCATCAAGAAGAAGACAGCACCAAGTGGCGCAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGGTGAATCTCGCGAGCTGAACACAGCGCACCCAGCATCTTCGGAGGTGAGCTGGCG 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCGCCACCGCGCGCTGAAGAAGAAGAAGAGCGTGACCGTGTGGAGCGTGGCGAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTTTCAGCGTGGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTrpThrAlaPheThrIlePro 300
QY 1072 AGCATCAACACAGACACCCCGGCGATCCGCTACAGTACAACTGTGTCGCCCGCAGGCGTG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGGAGCGCCAGCATCTTCCAGAGCAGCATCACCAAGATCTGGAGCGCTTCCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPhePheLysLys 340
QY 1192 CGCAACCCCGAGATCGTGATCTACCCAG-----GCCCGCCCTGTACGTGGCGACGACCTG 1245
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
QY 1246 GAGATGGCGCGACACCGCGCGCAAGATCGAGAGCTGGCGAAGCACCTGTGCGCTGGCGC 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380
```

RESULT 5

```
QY 1306 TTCACACCCCGCCACAAGAAGCACCAAGAAGAGCCCCCTTCTCTGGATGGGCTACGAG 1365
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1366 CTGCACCCCGCACAAAGTGGACCGTGGAGCCCATCGAGCTGCCCGAGAGAGAGAGCTGGACC 1425
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1426 GTCAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCGCCAGCAGATCTACCCCGGC 1485
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
QY 1486 ATCAAGGTGGCGCGCATGCTGCTCGCGCGCGCCCAAGGCCCTGACCGACATCGTG 1545
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1546 CCCTGACCGAGAGGCGAGCTGGAGCTGGCGCGAGAAACCGGAGATCTCTGCCGAGGCC 1605
Db 461 ProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1606 GTGCACCGGTGTACTACGACCCCGCAGGACCTGTGGCGGAGATCCAGAGAGCGGCC 1665
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1666 CACGACAGTGGACCTTACCAGATCTACACGAGGACCTTCAAGAACCTGAAGACCGGCAAG 1725
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1726 TAGCCCAAGATGGCGACCCCGCCACCAACGACGCTGAAGAGCTGACCGAGCGCGTGCAG 1785
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1786 AAGATCGCCATCGAGAGCATCGTGACTCGGCGCAAGACCCCAAGTTCGCGCTGCCCATC 1845
Db 541 LysIleThrThrGlnSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1846 CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGCGCCACCTGGATCCCCGAG 1905
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1906 TGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTGTACACAGCTGGAGAGAGGCC 1965
Db 581 TrpGluPheValAsnThrProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1966 ATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCAGACCAAGATCGGC 2025
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2026 AAGCCCGGCTTACGTGACCCGCGCGCGCGCAGAAATCGTGAGCCTGACCGAGACACC 2085
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2086 AACCAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTCGCAGGACACGCGGACGAGGTG 2145
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2146 AACATGCTCACCGACAGCAGTACGCGCTGGCGCATCATCCAGCCCGCAGCCCGACAGAGC 2205
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2206 GAGAGCGAGCTGTGAACACGAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTG 2265
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2266 AGCTGGGTGCGCGCCCAAGGCGATCGCGCGCAACGAGCAGATCCAGAGCTGGTGAGC 2325
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2326 AAGGGCATCCGCAAGTGTGTTCTCTGGACGCGCATCGAT 2364
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

US-10-093-953A-26

; Sequence 26, Application US/10093953A

; Publication No. US20040105871A1

; GENERAL INFORMATION:

; APPLICANT: Robinson, Harriet L.

; APPLICANT: Smith, James M.

; APPLICANT: Hua, Jian

; APPLICANT: Moss, Bernard

; APPLICANT: Amara, Rama

; APPLICANT: Wyatt, Linda

; APPLICANT: Earl, Patricia

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING

; FILE REFERENCE: 12804-005002

; CURRENT APPLICATION NUMBER: US/10/093.953A

; CURRENT FILING DATE: 2002-03-08

; PRIOR FILING DATE: 2000-03-02

; PRIOR FILING DATE: 2000-03-02

; PRIOR FILING DATE: 2000-12-01

; PRIOR APPLICATION NUMBER: US 09/798,675

; PRIOR APPLICATION NUMBER: PCT/US01/06795

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: US 60/324,845

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: US 60/325,004

; PRIOR FILING DATE: 2001-09-26

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 26

; LENGTH: 739

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: protein encoded by construct of vaccine vector

; OTHER INFORMATION: pGA2 and insert JS2 expressing clade HIV-1 VL

US-10-093-953A-26

Alignment Scores:

Pred. No.: 8,24e-160 Length: 739

Score: 3452.00 Matches: 651

Percent Similarity: 93.45% Conservative: 34

Best Local Similarity: 88.81% Mismatches: 30

Query Match: 75.49% Indels: 18

DB: 16 Gaps: 4

US-09-610-313B-31 (1-2463) x US-10-093-953A-26 (1-739)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGCAGCGAGCAG 279

DB 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyAlaArgGluPheSerSerGluGln 20

QY 280 AACCGC-----GCGCAAGCCCGCGAGTTCCCGCAGCGAGCAGC 303

DB 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40

QY 304 CGCGAGCTGCGAGTG-----CGCGGCGACAACCCCGCAGCGAGCCGCGCGCGAGCGC 357

DB 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60

QY 358 CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGGTGAGC 411

DB 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80

QY 412 ATCAAGTGGCGCGCGAGTCAAGAGCGCTGTGTCAGACCGCGCGCGCGAGCACCGTG 471

DB 81 IleLeuIleGlyGlyGlnLeuLeuGluAlaLeuLeuAspThrGlyAlaAspThrVal 100

QY 472 CTGAGGAGATGACCTGCGCGCGAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGC 531

DB 101 LeuGluGluMetSerLeuProGlyArgTrpArgTrpProLeuMetIleGlyGlyIleGlyGly 120

QY 532 TTCAATCAAGGTGCGCCAGTAGCAGATCTCTGATCGAGATCTCGGCGCAAGAGCCCATC 591

DB 121 PheIleLeuValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLeuAlaIle 140

QY 592 GGCACCGTGTGATCGCGCCCGCCAGCGTGAACATCATCGCGCGCAACATCTGACCCAG 651

DB 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160

QY 652 CTGGCTGCACCCCTGAATTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAG 711

DB 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLeuLeuLys 180

QY 712 CCGCGCATGACCGCCCGCCAGGTGAACAGTGGCCCTCGACCGAGGAGAGATCAAGGCC 771

DB 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysAla 200

QY 772 CTGACCGCCATCTCGAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGGCGCCGAG 831

DB 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220

QY 832 AACCCCTTACCAACACCCCGCTGTTCGCCATCAAGAAGAGGACAGCACCAAGTGGCGAAG 891

DB 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

QY 892 CTGGTGGACTTCGCGAGCTGAACAGCGCACCCAGCATCTTCGGAGGTGCAGCTGGGC 951

DB 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260

QY 952 ATCCCGCCCGCGCGCTGGAAGAAGAGAGCTGACCGCTGCGAGCTGGCGGAC 1011

DB 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280

QY 1012 GCCTACTTTCAGCGTGCCTGCGAGGAGCTTCGCGAAGTACACCGCTTCACCATCCCC 1071

DB 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300

QY 1072 AGCATCAACAGAGAGCCCGCGATCCGCTACAGTACAGCTGCTGCCCGCGAGGTGG 1131

DB 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320

QY 1132 AAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCGGAGCCCTCCGCGCC 1191

DB 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340

QY 1192 CGCAACCCCGAGATCGTGATCTACCG-----GCCCGCTGTGCTGGCGAGCGACCTG 1245

DB 341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360

QY 1246 GAGATCGCGCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCTGGTGGGC 1305

DB 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380

QY 1306 TTCAACACCCCGCAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1365

DB 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400

QY 1366 CTGCACCCCGCAGAGTGGACCGCTGCGAGCCCATCGAGCTGCGCGAGAGAGAGAGTGGAC 1425

DB 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420

QY 1426 GTGAACGACATCAAGAGCTGGTGGGCAAGCTGAACCTGGGCGAGCGAGATCTACCCGCG 1485

DB 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440

QY 1486 ATCAAGTGGCGCGAGTGTGCAAGCTGCTGCGCGCGCGCGAGCGCTCGACCATCGTG 1545

DB 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460

QY 1546 CCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGCGAGCGCC 1605

DB 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480

QY 1606 GTGCACGGCGTGTACTACGACCCCGCAGGACCTGCTGGCGCGAGATCCAGAGACGGGC 1665

```
Db 481 ValHisGlyValTyrTyAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1666 CACGACAGTGGACCTACAGACTACACAGAGCCCTTCAAGAACCTGAAGACGGCAAG 1725
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1726 TAGCCAAAGTGGCCAGCCGCCACCAACACGAGCTGAAGCAGCTGACCGAGGCGGTGCAG 1785
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
QY 1786 AAGATCGCCATGGAGACATCGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATC 1845
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1846 CAGAAGAGACCTGGGAGACCTGTGTGACCGACCTACTGGCAGGCGACCTGATCCCGAG 1905
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1906 TGGAGTTCTGTGAACACCCCGCCCTGCTGAAGCTGTGTACAGCTGGAGAGGAGCC 1965
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTyrTrpGlnLeuGluLysGluPro 600
QY 1966 ATCATCGGCGCGAGACCTTCTACTGTGACGCGCGCCCAACCCCGAGACCAAGATCGG 2025
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2026 AAGCGCGCTACGTGACGACCGCGCGCGCGAGAGATCGTACGCTGACCGAGCACCC 2085
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2086 AACCAAGACCGAGCTGACGAGCCATCCAGCTGCGCCCTGCAGACACGCGCAGCGGTG 2145
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2146 AACATCGTGACGACGACCAAGTACGCGCTGGGATCATCCAGCGCCAGCCCGACAAGAG 2205
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2206 GAGAGCGAGCTGTGAACCAAGATCATCGACACCTGATCAAGAGGAGAGGTGTACCTG 2265
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2266 AGCTGGGTGCGCGCCACAGGCGCATCGCGGCAACGAGCAGATCGACAGCTGGTGAGC 2325
Db 701 AlaTrpValProAlaHisLysGlyIleGlyValAsnGluGlnValAspLysLeuValSer 720
QY 2326 AAGGCGATCCGCAAGGTGTCTTCTGACGCGCATCGAT 2364
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 6

```
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE OF INVENTION: AN IMMUNE RESPONSE
; FILE REFERENCE: 12804-005002
; CURRENT APPLICATION NUMBER: US/10/093,953A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
```

```
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein encoded by construct of vaccine vector
; OTHER INFORMATION: pGAL and vaccine insert expressing clade B HIV-1
; OTHER INFORMATION: Gag-Pol
; US-10-093-953A-31
```

```
Alignment Scores:
Pred. No.: 8,24e-160 Length: 739
Score: 3452.00 Matches: 651
Percent Similarity: 93.45% Conservative: 34
Best Local Similarity: 88.81% Mismatches: 30
Query Match: 75.49% Indels: 18
DB: 16 Gaps: 4
```

US-09-610-313B-31 (1-2463) x US-10-093-953A-31 (1-739)

```
QY 220 TTCTTCCGAGACCTGCGCTTCCCGAGGCAAGCCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerGluGln 20
QY 280 AACCGC-----GCCAACGCCACACG 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40
QY 304 CGCGAGCTCGAGGTG-----CGCGCGCAACCCCGCGAGCGCGCGCGCGCGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGGCACCCCTG-----AACTTCCCGAGATCACCTGTGGCAGCGCGCGCGCGCGCG 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGGTGGCGCGCGCGAGATCAAGGAGGCGCTGCTGGACACCGCGCGCGCGCGCGCG 471
Db 81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120
QY 532 TTCATCAAGGTGGCGCGCGAGTACGACAGATCTCTGATCGAGATCTGCGCGCAAGAGGCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCTGAACCTTCCCATCAGCCCATCGAGACCGCGCGCGCGCGCGCGCGCG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
QY 712 CCGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCGCCATCTGCGAGGAGATGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
```

QY 832 AACCCCTACACACCCCGGTGTTCCCATCAAGAGAGAGACACCAAGTGGCGCAAG 891  
DB 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGGTGGACTTCGCGAGCTGAACAACAGCGACCCAGGACTTCTGGAGGTGCAGCTGGC 951  
DB 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCACCCCGCGCCCTGAAGAAGAAGAGCGTGACCGTGTGGACGTGGCGAC 1011  
DB 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTTCAGCTGCGCCCTGGAGGACTTCCGCAAGTACACCGCTTCCACATCCCC 1071  
DB 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACAACAGACCCCGGCGATCCGCTACCAAGTACACCGCTGCGCCAGAGGTGG 1131  
DB 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGGAGCCCGACATCTTCCAGAGCAGCATGACCAAGATCTGGAGCCCTTCCGCGCC 1191  
DB 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCCGAGATCGTGATCTACCAG-----GCCCGCTGTACGTGGCGAGCACTG 1245  
DB 341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360  
QY 1246 GAGATCGGCGAGCAGCGCCAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGGC 1305  
DB 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380  
QY 1306 TTCACCAACCCCGACAGACACCAAGAGAGAGCCCGCTTCTGTGGATGGGCTACGAG 1365  
DB 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400  
QY 1366 CTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACC 1425  
DB 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1426 GTCAACGACATCCAGAAAGCTGTGGCAAGCTGAATGGCGCCAGCAGATCTACCCCGGC 1485  
DB 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440  
QY 1486 ATCAAGTGGCGCAGCTGTGCAAGCTGCTCGCGCGGCCCAAGCCCTTACCGACATCGT 1545  
DB 441 IleLysValArgGlnLeuGlyLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1546 CCCTGACCGAGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCC 1605  
DB 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLysGluPro 480  
QY 1606 GTCCAGCGGTGTACTACGACCCAGCAGAGCTGTGGCGCGAGATCCAGAGAGCGGC 1665  
DB 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1666 CACGACAGTGGACCTACCAAGATCTACCAAGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725  
DB 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1726 TAGCCCAAGATGGCGACCGCCACCAACAGCGTGAAGCAGTGCACCGAGCGCGTGCAG 1785  
DB 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540  
QY 1786 AGATCGCATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCTCCGCTCCCATC 1845  
DB 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1846 CAGAAGGAGACCTGGGAGACCTGTGTGACCGACTACTTGGCAGCGCCACCTGGATCCCGGAG 1905  
DB 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1906 TGGGAGTTCGTGAACACCCCGCTGTGTGAGCTGTGTGTACCGCTGGAGAGAGGCC 1965

DB 581 TrpGluPheValAsnThrProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600  
QY 1966 ATCATCGCGCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGC 2025  
DB 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2026 AAGCGCGCTGCTGACCGACCGCGCGCGGAGATCTGAGCCTGACCGAGACCAACC 2085  
DB 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2086 AACCAAGACCGAGCTGCGAGCCATCCAGCTGCGCTGCGAGGACGCGCGAGCGGTG 2145  
DB 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2146 AACATCTGTGACCGACCGCAGTACCGCTGGCATCTCCAGGCGCCAGCCGACCAAGAGC 2205  
DB 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2206 GAGAGCGAGCTGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGT 2265  
DB 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuLysLysGluLysValTyrLeu 700  
QY 2266 AGCTGGGTGCGCCCGCACAGGCGCATCGCGGCAACGAGCAGATCGACCAAGCTGTGAGC 2325  
DB 701 AlaTrpValProAlaHisLysGlyIleGlyLysGlnGlnValAspLysLeuValSer 720  
QY 2326 AAGCGCATCGCAAGGTGCTTCTGCGAGCGATCGAT 2364  
DB 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7  
US-10-325-468-23  
; Sequence 23, Application US/10325468  
; Publication No. US20040101823A1  
; GENERAL INFORMATION:  
; APPLICANT: Soong, Nay Wei  
; APPLICANT: Pekrun, Katja  
; APPLICANT: Shibata, Riri  
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL  
; FILE REFERENCE: 0166.210US  
; CURRENT APPLICATION NUMBER: US/10/325,468  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US 60/343,524  
; PRIOR FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 1003  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.10 protein Pol  
US-10-325-468-23

Alignment Scores:  
Pred. No.: 9,85e-159 Length: 1003  
Score: 3430.00 Matches: 638  
Percent Similarity: 94.73% Conservative: 45  
Best Local Similarity: 88.49% Mismatches: 32  
Query Match: 75.01% Indels: 6  
DB: 16 Gaps: 3

US-09-610-313B-31 (1-2463) x US-10-325-468-23 (1-1003)

QY 220 TTCTTCCGAGAGACTTGGCTTCCCGCCAGGCAAGCCCGCGAGTTCGCCAGCGAGCAG 279  
DB 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20  
QY 280 AACCGCGCCACAGCCCGCCAGCCCGCGAGCTGCGAGTGG-----CCGCGCGACACCCC 333  
DB 21 ThrArgAlaAsnSerProIleArgGluGlnValTrpArgArgAspAsnAsnSer 40

QY 334 CGCAGCGAGCCGCGCCGAGCCCGAGCCGACCTG-----AAGTTCCTCCCGAGATCACC 387  
 Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
 QY 388 CTGTGGCAGCCGCGCCCTGTGTGAGCATCAAGGTGGCGCCGAGATCAAGAGAGCCCTGCTG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleLysAlaGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCGCGCCGAGCAGACCCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAGACCC 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCCAGTACGACACGATCCTGATC 567  
 Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleProIle 120  
 QY 568 GAGATCTGCGCGCAAGAGCGCATCGGCACCGTGTGATCGCGCCCGCCCGCGTGAACATC 627  
 Db 121 AspileCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGCGCGCAATGTGTACCCAGCTGGGTGTCACCTGAACTTCCCATCAGCCCATC 687  
 Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACCGTCCCGTGAAGCTGAAGCCCGGATGAGCGCCCGCCAGGTGAAGCATGCGCC 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGGAGAGAAAGTCAAGGCGCTGACCGCATCTGCGAGGAGATGGAAGAGGAGGC 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 ARGATCACCAAGATCGCCCGGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAG 867  
 Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
 QY 868 AAGGACAGCACCAAGTGGCGCAGCTGGTGGACTTCGCGAGCTGACACGCGCACCCAG 927  
 Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240  
 QY 928 GACTTCTGGAGGTGACGTGGGCGATCCCCACCCCGCGCTGGAAGAGAGAGAGAGC 987  
 Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGACGTGGCGCAGCCTACTTACGCGTGCCTGAGCAGGAGCTTCCGC 1047  
 Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
 QY 1048 AAGTACACCCGCTTCCATCCCGAGCATCAACAGCAGACCCCGCGCATCCGCTACCG 1107  
 Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln 300  
 QY 1108 TACAACGTGTGCCCCAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167  
 Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 QY 1168 AAGATCCTGGAGCCCTTCCCGCGCCCGCAACCCCGAGATCGTGTACTACCGAG-----GCC 1221  
 Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleTyrGlnTyrMetAsp 340  
 QY 1222 CCCTGTACGTGGCGCAGCTGGAGATCGCGCAGCAGCCCGCGCATCGAGGAGCTG 1281  
 Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 QY 1282 CGCAAGCACCTGTGCTGGGCTTCCACACCCCGCCAGCAAGAGACACCAAGAGGAGCC 1341  
 Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCTGTGGATGGCTACGAGCTGACCCCGAGCAAGTGGACCGTGGAGCCCATCGAG 1401  
 Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

QY 1402 CTGCCCGCAGAGGAGAGCTGGACCCGTGAACACATCAGCAAGCTGGTGGCGAAGCTCAAC 1461  
 Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
 QY 1462 TGGCGCAGCCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGGC 1521  
 Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440  
 QY 1522 GCCAAGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCGCAGCTGGAGCTGGCGCGAG 1581  
 Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGlnLeuGluLeuAlaGlu 460  
 QY 1582 AACCGCGAGATCTGCGCGAGCCGCTGTACTACACCCCGCCAGAGAGAGCCTG 1641  
 Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480  
 QY 1642 GTGGCGGAGATCCAGAAAGGCGCCAGCACCTAGTGGACCTTACCAGATCTTACCAGAGGCC 1701  
 Db 481 IleValGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIlePheGlnGluPro 500  
 QY 1702 TTCAAGACCTGAAGACCGGCAAGTACGCAAGATCGGCACCGCCACACCAAGAGCGCTG 1761  
 Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgSerAlaHisThrAsnAspVal 520  
 QY 1762 AAGCAGCTGACCGAGGCGGTGCAAGAGATCGCCATGAGAGCATCGTATCTGGGGCAAG 1821  
 Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540  
 QY 1822 ACCCCCAAGTTCGCTGCTCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881  
 Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560  
 QY 1882 TGGCAGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGGAGCTG 1941  
 Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 580  
 QY 1942 TGTGTACAGCTGGAAGAGGAGCCATCATCGCGCGCGAGACCTTCTACGTGGACCGCGCC 2001  
 Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
 QY 2002 GCCAAGCCGCGAGACCAAGATCGCAAGGCGCGTACGTGACCCAGCCGCGCGCGAGAG 2061  
 Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620  
 QY 2062 ATCGTGTGCTGACCGAGACCCACCAAGACCGAGCTGAGCGCATCCAGCTGGCC 2121  
 Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
 QY 2122 CTCGAGCAGACCGCGCAGCGAGGTGAACATCGTGACCGCAGCCAGTACGCTGGGCATC 2181  
 Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
 QY 2182 ATCCAGGCCCGAGCCGCAAGAGCAGAGAGCTGGTGAACCCAGATCATCGAGAGCTG 2241  
 Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
 QY 2242 ATCAAGAGAGAGAGTGTACTACCTGAGCTGGTGGTCCCGCCCGCCAGAGGCGCATCGCGGCAAC 2301  
 Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
 QY 2302 GAGCAGATCCAGCAAGCTGGTGAAGGCGCATCCGCAAGGTGTGTCTTGGAGCGGCATC 2361  
 Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgValLeuPheLeuAspGlyIle 720  
 QY 2362 GAT 2364  
 Db 721 Glu 721

RESULT 8  
 US-10-325-468-35  
 ; Sequence 35, Application US/10325468  
 ; Publication No. US20040101823A1  
 ; GENERAL INFORMATION:

; APPLICANT: Soong, Nay Wei  
 ; APPLICANT: Pekrun, Katja  
 ; APPLICANT: Shibata, Riri  
 ; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL  
 ; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS  
 ; FILE REFERENCE: 0166.210US  
 ; CURRENT APPLICATION NUMBER: US/10/325,468  
 ; CURRENT FILING DATE: 2002-12-19  
 ; PRIOR APPLICATION NUMBER: US 60/343,524  
 ; PRIOR FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 56  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 35  
 ; LENGTH: 1003  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol  
 ; US-10-325-468-35

Alignment Scores:  
 Pred. No.: 1,23e-158 Length: 1003  
 Score: 3428.00 Matches: 639  
 Percent Similarity: 94.73% Conservative: 44  
 Best Local Similarity: 88.63% Mismatches: 32  
 Query Match: 74.96% Indels: 6  
 DB: 16 Gaps: 4

US-09-610-313B-31 (1-2463) x US-10-325-468-35 (1-1003)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGGCAGGCGCGAGTTCCCGAGGAGCAG 279  
 DB 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20  
 QY 280 AACCGCGCCAAACGCCACCAGCGCGAGCTGCAGGTG---CGCGCGAC---AACCCC 333  
 DB 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgGlyAspAsnAsnSer 40  
 QY 334 CGCAGCAGGCGCGCGCGAGCGCGAGCGCGACCCCTG-----AACTTCCCGCAGATCAC 387  
 DB 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
 QY 388 CTGTGGCAGGCGCCCTGCTGTGAGATCATAGTGGCGCGCGCGAGATCAAGAGGCGCTCTG 447  
 DB 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCGCGCGCAGCAGCACCGTCTCGAGGAGATGAGCTGCCCGCGCAAGTGGAAGCCC 507  
 DB 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCGGCGGCATCGCGCGCTTCATCAAGGTGGCGCGAGTACGACGACGATCCTGATC 567  
 DB 101 LysMetIleGlyIleGlyGlyPheIleLysValargGlnTyArgGlnIleProIle 120  
 QY 568 GAGATCTGGCGCAAGAGGCATCGGACCGTGTGATCGGCGCGCGCGCGCGCGCGCGCGCG 627  
 DB 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGGCGCAACATGCTGACCCAGCTGGCTGACCTGCACTTCCCATCAGCCCGCGCGCG 687  
 DB 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACCGTGGCGTGAAGCTGAAGCGCGCGATGACGCGCGCGCGCGCGCGCGCGCGCGCG 747  
 DB 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGAGAGAGATCAGGCGCGCTGACCGCGCATCTGCGAGGAGATGAGAGAGGCGG 807  
 DB 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCACCAAGATCGGCGCGCGAGAACCCCTTACAAACACCCCGCGTTCGCGCATCAAGA 867  
 DB 201 LysIleSerLysIleGlyProGluAsnProTyArgAsnThrProValPheAlaIleLysLys 220

QY 868 AAGCAGACCAAGTGGCGCAAGCTGGTGAGCTTCCGCGAGCTGAAACAAGCGCACCCAG 927  
 DB 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240  
 QY 928 GACTTCTGGGAGGTGACGTGGGCATCCCGACCCCGCGCGCTCAAGAAGAAGAGAGC 987  
 DB 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGGACGTGGCGCAGCCCTTACCTTACGCGTCCCGCGAGAGAGACTTCGCG 1047  
 DB 261 ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheArg 280  
 QY 1048 AAGTACACCGCTTTCATCCCGCAGCATCAACACGAGACCCCGCGCATCCGCTTACCG 1107  
 DB 281 LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln 300  
 QY 1108 TACAACCTGTGCGCCAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167  
 DB 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 QY 1168 AAGATCTGGAGCGCTTCCGCGCGCCCAACCCCGAGATCGTGATCTACCAG-----GCC 1221  
 DB 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyGlnTyMetAsp 340  
 QY 1222 CCCCTGTACGTGGCGCAGCACCTGGAGATCGCGCAGCACCGCGCAAGATCGAGAGCTG 1281  
 DB 341 AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 QY 1282 CGCAAGCACCTGTGCGCTGGCGCTTCCACCAACCCCGAGAGAGACCCCAAGAGAGAGCC 1341  
 DB 361 ArgGlnHisLeuLeuTyTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCTGTGTGGGTACGAGCTGCACCCCGACCAAGTGGACCGTGCAGCCCATCCAG 1401  
 DB 381 ProPheLeuTrpMetGlyTyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
 QY 1402 CTGCGCGAAGAGAGAGTGGACCGGTGAAGACATCCAGAACGCTGGTGGGCAAGCTGAAC 1461  
 DB 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
 QY 1462 TGGCGCAGCAGATCTACCCCGCGCATCAAGTGGCGCGAGTGTGCAGAGCTCTCGCGGC 1521  
 DB 421 TrpAlaSerGlnIleTyAlaGlyIleLysValLysGlnLeuLysLysLeuLysGly 440  
 QY 1522 GCCAAGCCCTGACCGACATCGTCCCTCGACCGAGGCGCGAGCTGGAGCTGGCCGAG 1581  
 DB 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuLysAlaGlu 460  
 QY 1582 AACCGCAGATCTCGCGCAGCGCGTGCAGCGGTGTACTACGACCCCGCAGCAAGAGCTG 1641  
 DB 461 AsnArgGluIleLeuLysGluProValHisGlyValTyArgProSerLysAspLeu 480  
 QY 1642 GTGGCGCAGATCCAGAGCAGGCGCCACGACCGATGGAGCTTACAGATCTACCGAGCCG 1701  
 DB 481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyGlnIlePheGlnGluPro 500  
 QY 1702 TTCAAGAACCTGAAGACCGCGCAAGTACGCCAAGATGGCGCGCGCGCGCGCGCGCGCG 1761  
 DB 501 PheLysAsnLeuLysThrGlyLysTyArgLysThrArgGlyAlaHisThrAsnAspVal 520  
 QY 1762 AAGCAGCTGACCGAGCGCGTGCAGAAAGATCCCATGAGAGCATCTGTGATCTGGGCAAG 1821  
 DB 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540  
 QY 1822 ACCCGCAAGTTCGCTGCGCCATCCAGAGAGACTGGGAGACCTGGTGCAGCGACTAC 1881  
 DB 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTy 560  
 QY 1882 TGGCAGGCGCACCTGGATCCCGCGAGTGGAGTTCGTGAACACCCCGCGCTGGTGAAGCTG 1941  
 DB 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580



```
QY 1942 TGGTACCAGCTGGAGAGAGCCATCATCGGCGCGGAGACCTTCTACGTGGACGGCGCC 2001
Db 581 TTPTyrglnleuGluLysGluProlleValGlyAlaGluThrPheTyValAspGlyAla 600
QY 2002 GCCAACCGCGAGACCAAGATCGGCAAGGCGCGGTACGTGACCGACCGGGCGGCGAGAG 2061
Db 601 AlaAsnArggluthrLysLeuGlyLysAlaGlyTyValThrSerArgGlyArgGlnLys 620
QY 2062 ATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGGTGAGGCCATCCAGCTGGCC 2121
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2122 CTGCAGGACAGCGCGCGAGGTCAACATCGTCACCGACAGCAGGTACGCGCTGGGCATC 2181
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIle 660
QY 2182 ATCCAGGCCAGCCGACCAAGAGCGAGAGAGCTGGTGAACAGATCATCGAGCGAGCTG 2241
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGNAGGAGAGGTGTACCTGAGCTGGTGGTGGCCGCCACAGGCGCATCGGGCGAAC 2301
Db 681 IleLysLysGlnLysValTyLeuThrTrpIleProAlaHisLysGlyIleGlyGlyAsn 700
QY 2302 GAGCAGATCAGCAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTTGGACGGCATC 2361
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
Db 721 Glu 721
```

## RESULT 9

```
US-10-325-468-9
; Sequence 9, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Shibata, Riri
; APPLICANT: Pekrun, Katja
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;
; OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;
; OTHER INFORMATION: clone P10.26 protein Pol
US-10-325-468-9
```

## Alignment Scores:

Pred. No.:	1.38e-158	Length:	1003
Score:	3427.00	Matches:	638
Percent Similarity:	94.59%	Conservative:	44
Best Local Similarity:	88.49%	Mismatches:	33
Query Match:	74.94%	Indels:	6
DB:	16	Gaps:	3

US-09-610-313B-31 (1-2463) x US-10-325-468-9 (1-1003)

```
QY 220 TTCTTCGGCAGGACCTGGCTTCCCGCAGGGNAGCGCCGCGAGTTCCTCCCGAGGACGAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
```

```
QY 280 AACCGCGCAACAGCCCGCCACCGCCAGCTGCAGGTG-----CGCGCGCAACAACCCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
QY 334 CGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGTGACGCGCCCTGTGTGAGCATCAAGGTGGGGCGGCAGATCAAGGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
QY 508 AGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCGCGCGCGCGCGCGCGCGCG 567
Db 101 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProIle 120
QY 568 GAGATCTCGCGCAAGAGCCATCGCGCACCGCTGTGATCGGCCCGCCACCGCGCGCGCG 627
Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGGTGCACCTGAACCTTCCCATCAGCCCGCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGAACTGAAGCCCGCGCATGGAGCGCGCGCGCGCGCGCGCGCGCGCG 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAAAGAGGGCG 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240
QY 928 GACTTCTGGGAGGTGCAGCTGGGCATCCCGCCACCGCGCGCGCGCGCGCGCGCGCGCG 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGACGTGGGCGACGCGCTACTTCAGCGTGGCCCTTGGACGAGGACTTCGCG 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCTTACCATCCCGCGCATCAACAGGAGACCGCGCGCGCGCGCGCGCGCGCG 1107
Db 281 LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln 300
QY 1108 TACAACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCTGTGAGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1221
Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyGlnTyMetAsp 340
QY 1222 CCCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1281
Db 341 AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 CGCAAGCACCTGTGTGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1341
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCCTTCTGTGTGAGTGGGCTACGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1401
```

```

Db      381  ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
Qy      1402 CTGCCCGAGAGAGAGCTGGACCGTGAACACATCCAGAACCTGGTGGGCAAGCTCAAC 1461
Db      401  LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
Qy      1462 TGGCCAGCAGATCTACCCCGGATCAAGTGGCCAGCTGTCAGAGCTGCTGCGGGC 1521
Db      421  TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
Qy      1522 GCCAAGSCCTGACCGACATGTCGCCCTGACCGAGGAGCGGAGCTGGAGCTGGCCGAG 1581
Db      441  ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460
Qy      1582 AACCGCAGATCTTCGCGAGCCCGCTGACGGCTGTACTACGACCCCGACGAGCCTG 1641
Db      461  AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
Qy      1642 GTGCCCGAGATCCAGAGAGCGGCGACGACAGTGGACCTACCAATCTACAGAGGCC 1701
Db      481  IleValGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500
Qy      1702 TTCAGAACCTGAGACCGGCAAGTACGCCAAGATGCGCACCGCCGACACCAAGAGCTG 1761
Db      501  PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520
Qy      1762 AAGCAGCTGACCGAGGCGGTGCAGAGATCCGATGAGAGCATCGTATCTGGGGCAAG 1821
Db      521  LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
Qy      1822 ACCCCAAAGTTCGCCCTGCCATCCAGAGAGACCTGGGAGACCTGGTGACCGACTAC 1881
Db      541  IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr 560
Qy      1882 TGGCAGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTCGTGAAGCTG 1941
Db      561  TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
Qy      1942 TCGTACAGCTGGAGAGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAGCGGCC 2001
Db      581  TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
Qy      2002 GCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGAGCCCGCGCAGAAG 2061
Db      601  AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620
Qy      2062 ATCGTAGCTGACCGAGACCAACCAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
Db      621  ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
Qy      2122 CTGCAGGACAGCGCAGCAGAGGTGAACATCGTACCGACGACAGCCAGTACGCCCTGGGCATC 2181
Db      641  LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Qy      2182 ATCCAGCCCGCAGCCGCAAGAGCAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTG 2241
Db      661  IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
Qy      2242 ATCAAGAGAGAGAGTGTACTGAGCTGGGTGCCCGCCACAGGGCATCGCGGGCAAC 2301
Db      681  IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
Qy      2302 GAGCAGATCCACAGCTGGTGGAGCAGGGCATCCGCAAGGTGCTGTCTCGGAGCGGCATC 2361
Db      701  GluGlnValAspLysLeuValSerAlaGlyIleAargArgValLeuPheLeuAspGlyIle 720
Qy      2362 GAT 2364
Db      721  Glu 721

```

RESULT 10  
US-10-325-468-20

```

; Sequence 20, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20

Alignment Scores:
Pred. No.: 1.38e-158 Length: 1003
Score: 3427.00 Matches: 638
Percent Similarity: 94.59% Conservative: 44
Best Local Similarity: 88.49% Mismatches: 33
Query Match: 74.94% Indels: 6
DB: 16 Gaps: 3

US-09-610-313B-31 (1-2463) x US-10-325-468-20 (1-1003)
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
Qy 280 AACCGCGCCAAACGCCCGCCAGCCAGCGAGCTGCAGGTG-----CCGCGGCAACACCCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
Qy 334 CGCAGCGAGCGCGCGCGCGAGCCCGAGCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCG 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
Qy 388 CTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
Qy 448 GACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
Qy 508 AACATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
Db 101 LysMetIleGlyIleGlyIleGlyIleLysValArgGlnTyrAspGlnIleProIle 120
Qy 568 GAGATCTCGCGCAAGAGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
Qy 628 ATCGCGCGCGCAACATGTGTGACCCAGCTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
Qy 688 GAGACCGTGCCTGTGAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
Qy 748 CTGACCGAGGAGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

```

```
QY 808 AAGATCAACCAAGATCGGCCCCGAGAACCCCTCAACAACACCCCGTGTTCGCCCATCAAGAAG 867
Db 201 LysileSerLysileGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGCAGACCAAGTGGCCCAAGCTGTGGTGAATTCCTGGGAGCTGAACAAGCCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240
QY 928 GACTTCTGGGAGGTGGAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAGAAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLys 260
QY 988 GTGACCGCTGTGACGTGGCGCAGCGCTACTTCAGGGTGGCCCTGGACGAGGACTTCGCG 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCCTTCCATCCCGAGCATCAACAACGAGACCCCGGCGATCCGCTACCAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAAGCTGTGCTCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCCTGGAGCCCTTCCCGCGCCCGCAACCCCGAGATCGTGATCTACCAG-----GCC 1221
Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340
QY 1222 CCCTGTACTGTGGCGCAGCACTGGAGATCGGCCAGCACCGCCCAAGATCGAGGAGCTG 1281
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 CGCAAGCACTGTGCGCTGGCGCTTCCACCACCCCGCAGCAAGAAGACCAAGAGGAGGCC 1341
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCTTCTCTGTGGTGGCTACGAGCTGCACCCGCAAGTGGACCGTGCAGCCCATCGAG 1401
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1402 CTGCCCGAGAAGAGAGCTGGACCGTGAACGATTCAGAAAGCTGTGGCAAGCTGAAC 1461
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGGCGAGCCAGATCTACCCCGCCATCAAGTGCAGCTGCGCAAGCTGCTGCGCGGC 1521
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
QY 1522 GCCAAGGCCCTGACCGACATCGTCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAG 1581
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460
QY 1582 AACCGGAGATCTTGGCGAGCCGCTGCACGGGTGTACTACGACCCCGCAGCAAGACCTG 1641
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1642 GTGGCCGAGATCCAGAGAGGCGCCAGCACGATGGACCTTACCAGATCTTACAGAGGCC 1701
Db 481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTrpGlnIlePheGlnGluPro 500
QY 1702 TTCAAGAACCTGAAGACCGCAAGTACGCCAAGATGCGCACCCCGCCACACCAACGACGTG 1761
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520
QY 1762 AAGCAGCTGACCGAGGCGGTGCGAAGATCCCATGGAGAGCATCGTGATCTGGGGCAAG 1821
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
QY 1822 ACCCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGCGACCGACTAC 1881
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr 560
QY 1882 TGGCAGGCCACCTGGATCCCGGAGTGGGAGTTCTGTGAACACCCCGCCCTGTGTGAAGCTG 1941
```

```
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1942 TGGTACCACTCGAGAAAGAGCCCATCATCGCGCGGAGACCTTCTACCTGAGCGCGGCC 2001
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCGCGAGACCAAGATCGGCAAGCGCGGTCTACGTACCCGAGACCGGGCGCGCAGAAG 2061
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620
QY 2062 ATCTGTAGCTGACCCGAGACCAACCAACGAGAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2122 CTGCGAGCAGCGGCGAGCGGTGACATCGTGACCGAGCGAGCTACGCCCTGGGCGCATC 2181
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2182 ATCCAGGCCCGACCCGACCAAGAGCGAGCGAGCTGGTGAACCCAGATCATTCGAGCAGCTG 2241
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAGGAGAGGTGTACTGTAGCTGGGTGCGCGCCGACCAAGGGCATCGCGCGCAAC 2301
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2302 GAGCAGATCGACNAAGCTGTGTGACGAGGCGCATCCGCAAGCTGTCTTCTGGACGCGCATC 2361
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
Db 721 Glu 721

RESULT 11
US-10-296-734-2
; Sequence 2, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT FILING DATE: 2003-08-04
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 995
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: POL consensus polypeptide
US-10-296-734-2

Alignment Scores:
Pred. No.: 3.37e-158 Length: 995
Score: 3419.00 Matches: 640
Percent Similarity: 94.44% Conservative: 40
Best Local Similarity: 88.89% Mismatches: 28
Query Match: 74.76% Indels: 12
DB: 15 Gaps: 4

US-09-610-313B-31 (1-2463) x US-10-296-734-2 (1-995)
QY 220 TTCTTCCGGAGGAGCTGGCTTCCCGCAGGCGCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyLysAlaArgGluPheSerGluGln 20
QY 280 AACCGCGCCAACAGCCCGCCAGCTGCGAGTGCAGCGCGCGCGAGCAACCCCGCAGC 339
```

Db 21 ThrGlyAlaAsnSerSerAlaSerArgLysLeu-----GlyAspGly----- 34  
QY 340 GAGCGCGCGCGAGCGCCAGGGCACC-----CTGAATCTCCCCAGATCACCCTG 390  
Db 35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53  
QY 391 TGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGCCAGATCAAGAGAGCGCTGTGGAC 450  
Db 54 TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 73  
QY 451 ACCGGCGCCACGACACCGGTCTGGAGGAGATGAGCTGGCCCGCAAGTGGNAGCCCAAG 510  
Db 74 ThrGlyAlaAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys 93  
QY 511 ATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCCAGTACGACCATGCTGTATCGAG 570  
Db 94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnThrAspGlnIleLeuIleGlu 113  
QY 571 ATCTGCGGCAAGAGCGCATCGGCACCGCTGTGATCGCGCCCGCCCGCTGAACATCATC 630  
Db 114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133  
QY 631 GCGCGCAACATGCTGACCGCAGCTGGCTGGCTGACCTGACCTGACCTCCCATCAGCCCATCGAG 690  
Db 134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153  
QY 691 ACCGTGCGCTGAAGCTGAAGCGCGCATGACGCGCCCGCAAGGTGAAGCAGTGGCCCTG 750  
Db 154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173  
QY 751 ACCGAGGAGAGATCAAGCGCCCTGACCGCCATCTGCGAGGAGATGGAAGAGGGCAAG 810  
Db 174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGlyLys 193  
QY 811 ATACCAAGATCGCGCGCGAGAACCTCTACACACCGCGCTGTCGCGCATCAAGAGAAG 870  
Db 194 IleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys 213  
QY 871 GACAGCACCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTGAACAGCGCACCCAGGAC 930  
Db 214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233  
QY 931 TTCTGGAGGTGACAGCTGGCGCATCCCCACCGCCCGCTGAGAGAGAGAGAGCGTG 990  
Db 234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 253  
QY 991 ACCGTGCTGGAGCTGGCGGCGCTACTTACGTGCGCTGCGCGAGGAGCTTCCGCAAG 1050  
Db 254 ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluSerPheArgLys 273  
QY 1051 TACACCGCTTCACCATCCCGCATCAACAGAGAGACCGCGCATCCGCTACAGTAC 1110  
Db 274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGlnTrp 293  
QY 1111 AACGTGCTGCGCCAGGCTGGAAGGGCAGCGCCAGCATCTTCCAGAGCAGCATGACCAAG 1170  
Db 294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313  
QY 1171 ATCTGAGGCGCTTCCCGCGCGCAACCGCGAGATCGTATCTACCGAG-----GCCCGC 1224  
Db 314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyGlnTrpMetAspAsp 333  
QY 1225 CTGTACTGGCGCAGCATCGGAGATCGCGCAGCAGCGCCAGATCGAGGAGCTCGCG 1284  
Db 334 LeuTyThrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353  
QY 1285 AAGCACCTGCTGCGGCTTCCACCAACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1344  
Db 354 AlaHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373  
QY 1345 TTCTGTGGATGGCTACGAGCTGCACCGCGACAAAGTGGACCGGTGACCGCATCGAGCTG 1404

Db 374 PheLeuTrpMetGlyTyThrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393  
QY 1405 CCGAGAGAGAGAGCTGGACCGCTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACCTG 1464  
Db 394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413  
QY 1465 GCAGAGCAGATCTACCCCGCGCATCAAGGTGGCGCGCAGCTGTGCAAGCTGTGCGCGCGCC 1524  
Db 414 AlaSerGlnIleTyAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433  
QY 1525 AAGSCCTCTGACGACATCGTGGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAAC 1584  
Db 434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluGluAsn 453  
QY 1585 CCGAGATCTCTGGCGGAGCGCGTGCACGGGTGTACTACGACCCCGACGAGAGAGCTGGTG 1644  
Db 454 ArgGluIleLeuArgGluProValHisGlyValTyThrAspProSerLysAspLeuVal 473  
QY 1645 GCGAGATCCAGAGAGCGCGCGCAGCAGCAGTGGACCTACAGATCTACAGGAGCCCTTC 1704  
Db 474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyThrGlnIleTyThrGlnProPhe 493  
QY 1705 AAGAACCTGAAGACCGCAAGTACGCCAAGATGGCCACCGCCACCAACGAGCTGAAG 1764  
Db 494 LysAsnLeuLysThrGlyLysTyThrSerArgLysArgSerAlaHisThrAsnAspValArg 513  
QY 1765 CAGCTGACCGAGGCGGTGCAGAGATCGCCATCGAGAGCATCTGTGATCTGGGCAAGACC 1824  
Db 514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533  
QY 1825 CCAGAGTTCGCTGCGCCATCCAGAGAGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884  
Db 534 ProlLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyTrp 553  
QY 1885 CAGSCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGG 1944  
Db 554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573  
QY 1945 TACAGCTGAGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC 2004  
Db 574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyThrValAspGlyAlaAla 593  
QY 2005 AACCGGAGACCAAGATCGGAGCGCGCTAGCTGACCGAGCCGGCGCGCGAGAGATC 2064  
Db 594 SerArgGluThrLysLeuGlyLysAlaGlyTyThrValThrAspArgGlyArgGlnLysVal 613  
QY 2065 GTGAGCTGACCGGAGACCCACCAAGAGAGAGCTGAGGCGCATCCAGCTGGCGCTG 2124  
Db 614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633  
QY 2125 CAGGACAGCGCGCAGCGAGGTGAACATCGTGACCGACAGCGAGTACGCGCTGGGATCATC 2184  
Db 634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyThrAlaLeuGlyIleIle 653  
QY 2185 CAGCGCGCGCGCAGAGAGAGAGAGAGCTGGTGAACAGAGATCATCGAGAGAGCTGATC 2244  
Db 654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673  
QY 2245 AAGAGAGAGAGAGTGTACCTGAGCTGGTGGTCCCGCCACAGGAGGATCGCGCGCAAGCAG 2304  
Db 674 LysLysGlnLysValTyThrLeuSerTrpValProAlaHisLysGlyIleGlyLysGlu 693  
QY 2305 CAGATCCAGAGCTGGTGAAGAGGCGCATCCGCAAGGTGTCTTCTGGAGCGGATCGAT 2364  
Db 694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713

## RESULT 12

US-10-296-734-1470  
; Sequence 1470, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A



[illegible]



```

; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (335)..(335)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (365)..(365)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (369)..(369)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (373)..(373)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (402)..(403)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (434)..(434)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (439)..(439)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (452)..(452)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (470)..(470)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (480)..(480)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (487)..(487)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (514)..(514)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (534)..(535)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (544)..(544)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (553)..(553)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (593)..(593)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (610)..(610)
; OTHER INFORMATION: X = unknown amino acid

```

```

; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (675)..(675)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:

Alignment Scores:
Pred. No.: 4.62e-150 Length: 1006
Score: 3251.50 Matches: 618
Percent Similarity: 88.81% Conservatives: 25
Best Local Similarity: 85.36% Mismatches: 72
Query Match: 71.10% Indels: 9
DB: 15 Gaps: 3

US-09-610-313B-31 (1-2463) x US-10-296-734-1471 (1-1006)
QY 220 TTCTTCGCGAGGACCTGGCTTCCCTCCAGGGCAAGCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArg***PheSerSerGluGln 20
QY 280 AACCGCGCCCAACAGCCCCACAGCCGCGAGCTGCAGTGGCGCGGACCAACCC----- 333
Db 21 Thr*****AsnSer***ThrSerArg***LeuTrpAspGlyGly***Asp*****Leu 40
QY 334 CGCAGCGAGGCGCGCGCGCGAGCCAGCGCCAGGGC-----ACCTGTGAATTCCTCC 378
Db 41 *****Gly***Glu***GlnGly***Gly*****SerPhe***PhePro 60
QY 379 CAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGGCGCGAGTCAAGGAG 438
Db 61 GlnIleThrLeuTrpGlnArgProLeuValThrVal***Ile***GlyGlnLeuIleGlu 80
QY 439 GCCTGTCTGGACACCGCGCGCGAGCACCGCTGCTGGAGGAGATGACCTGCCCGGCAAG 498
Db 81 AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAsnLeuProGlyLys 100
QY 499 TGAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCAAGGTGGCGCGCGAGTCAAGCAG 558
Db 101 TrpLysProLys***IleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGln 120
QY 559 ATCTGTATCGAGATTCGCGCAAGAGCCATCGGCACCGTGTGTGATCGGCCCGCCACCCC 618
Db 121 IleLeuIleGluIleCysGlyLysLys***IleGlyThrValLeuValGlyProThrPro 140
QY 619 GTGAACATCATCGCGCGCAACATGCTGACCCGCTGGCTGCGACCTGCACTTCCCATC 678
Db 141 ValAsnIleIleGlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160
QY 679 AGCCCATCGAGACCCGTCGCGTGAAGCTGAAGCCGCGCATGCGCGCGCCCAAGGTGAAG 738
Db 161 SerProIleGluThrValProValLysLeuLysPro***MetAspGlyProLysValLys 180
QY 739 CAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCCATCTGCGAGGAGATGGAG 798
Db 181 GlnTrpProLeuThrGluGluLysIleLysAlaLeuThr***IleCys***GluMetGlu 200
QY 799 AAGGAGGCAAGATCACCAAGATCGCGCGCGAGAACCCCTACACACCCCGCTGTCGCC 858
Db 201 LysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAla 220
QY 859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCGCGCGAGCTGAACAAG 918
Db 221 IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 240
QY 919 CGCACCCAGGACTTCTGGGAGGTGCGAGTGGGGCATCCCCACCCCGCGCGCTGAAGAAG 978
Db 241 ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis***AlaGlyLeuLysLys 260
QY 979 AACAAGCGCTGACCGCTGCTGGACGTGGCGCGAGCCCTACTTCAAGCTGCCCTGGAGCAG 1038
Db 261 LysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAsp*** 280
QY 1039 GACTTCGCAAGTACACCGCCTTCACCATCCCGAGCATCAACACAGAGACCCCGCGCATC 1098

```

Db 281 \*\*\*PheArgLysTyrThrAlaPheThrIleProSer\*\*\*AsnAsnGluThrProGly\*\*\* 300  
Qy 1099 CGCTACCAAGTACAAAGTCTCCAGGCTGGAGGCGACCCAGCATCTTCCAGAGC 1158  
Db 301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro\*\*\*IlePheGln\*\*\* 320  
Qy 1159 AGCATGACCAAGATCTCTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCTACCAAG 1218  
Db 321 SerMetThrLysIleLeuGluProPheArg\*\*\*\*\*Pro\*\*\*IleValIleTyrGln 340  
Qy 1219 -----GCCCCCTGTACGTGGGAGCGACCTGAGATCGCCAGCAGCCGCGCAAGATC 1272  
Db 341 TyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360  
Qy 1273 GAGGAGCTGGCCCAAGCAGCTCTCGCTGGGGCTTCCACACCCCGACCAAGAACACACAG 1332  
Db 361 GluGluLeuArg\*\*\*HisLeuLeu\*\*\*TrpGlyPhe\*\*\*ThrProAspLysLysHisGln 380  
Qy 1333 AAGGAGCCCTCTCTGTGGATGGGTACAGCTGCACCCCGCAAGAGTGGACCGTGCAG 1392  
Db 381 LysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTrpThrValGln 400  
Qy 1393 CCCATCAGCTGCCGAGAGGAGCTGGACCTGACGACATCCAGACATCCAGAGCTGGTGGC 1452  
Db 401 Pro\*\*\*\*\*LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420  
Qy 1453 AAGCTGAACCTGGGCGCAGCATCTACCCCGCATCAAGGTGCCCGCAGCTGTGCAAGCTG 1512  
Db 421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys\*\*\*LysGlnLeuCys\*\*\*Leu 440  
Qy 1513 CTGCGCGCGCCCAAGGCCCTGACCGACATCGTGCCTCCCTGACCGAGGCGCGAGCTGAG 1572  
Db 441 LeuArgGlyAlaLysAlaLeuThrAspIleVal\*\*\*LeuThrGluGluAlaGluLeuGlu 460  
Qy 1573 CTGCGCGAGAACCGGAGATCTCGCGAGCCCTGCACCGCTGTACTAGCACCCCGCAGC 1632  
Db 461 LeuAlaGluAsnArgGluIleLeuLys\*\*\*ProValHisGlyValTyrAspPro\*\*\* 480  
Qy 1633 AAGGACCTGTGGCGGAGATCCAGAAGCAGGCGCCACACAGTGCAGCTCACCAGATCTAC 1692  
Db 481 LysAspLeuValAlaGlu\*\*\*GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500  
Qy 1693 CAGGAGCCTTCAAGAACCTGAAGACCGGAGTACGCCAAGATGCGCACCGCCACAC 1752  
Db 501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla\*\*\*LysArgSerAlaHisThr 520  
Qy 1753 ACGACGTGAGCAGCTGACGAGCGCGTGCAGAAAGATCGCCATGCGGAGCATCTGTATC 1812  
Db 521 AsnAspValLysGlnLeuThrGluValValGlnLysVal\*\*\*\*\*GluSerIleValIle 540  
Qy 1813 TGGGGCAAGACCCCAAGTTCCGCTCCCATCCAGAAGGAGACCTGGGAGACCTGTGTGG 1872  
Db 541 TrpGlyLys\*\*\*ProLysPheArgLeuProIleGln\*\*\*GluThrTrpGlu\*\*\*TrpTrp 560  
Qy 1873 ACCGACTACTGCGAGCGCACCTGTGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTG 1932  
Db 561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580  
Qy 1933 GTGAAGCTGTGTACCAAGCTGGAGAGAGGCCCATCATCTCGCGCCGAGACCTTCTAGCTG 1992  
Db 581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle\*\*\*GlyAlaGluThrPheTyrVal 600  
Qy 1993 GACGGCGCCCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGCGCGGC 2052  
Db 601 AspGlyAlaAlaAsnArgGluThrLys\*\*\*GlyLysAlaGlyTyrValThrAspArgGly 620  
Qy 2053 CGGCAAGATCTGTGACCTGACCGACACCAACAGAGACCGGAGCTGCAGCGCCATC 2112  
Db 621 ArgGlnLysValValSerLeuThrGluThrAsnGlnLysThrGluLeuHisAlaIle 640  
Qy 2113 CAGCTGCGCTTCAGGACAGCGGAGGTGAACATCTGTGACCGACCGCATGACGCC 2172

Db 641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660  
Qy 2173 CTGGGCTATCTCCAGGCGCCAGCCGACAGAGCGAGAGCGAGCTGGTGAACACGATCATC 2232  
Db 661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu\*\*\*ValAsnGlnIleIle 680  
Qy 2233 GAGCAGCTGATCAAGAAGGAGAGGTGTACTGAGCTGGGTGGCCGCCCAAGAGGGCATC 2292  
Db 681 GluLysLeuIle\*\*\*Lys\*\*\*LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700  
Qy 2293 GGCGGCAACGAGCAGATCCAGCAAGCTGTGAGCAAGGCGCATCCGCAAGGTCTGTCTCTG 2352  
Db 701 GlyGlyAsnGluGlnValAspLysLeuValSer\*\*\*GlyIleArgLysValLeuPheLeu 720  
Qy 2353 GACGCGATCGAT 2364  
Db 721 AspGlyIleAsp 724  
RESULT 14  
US-10-301-661A-6  
; Sequence 6, Application US/10301661A  
; Publication No. US20030157660A1  
; GENERAL INFORMATION:  
; APPLICANT: INSERM  
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: MAUCLERE, Philippe  
; APPLICANT: LOUSSERT-AJAKA, Ibtissem  
; APPLICANT: SIMON, Francois  
; APPLICANT: SARAGOSTI, Sentob  
; APPLICANT: BARRE-SINOSSI, Françoise  
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.  
; FILE REFERENCE: 598US12  
; CURRENT APPLICATION NUMBER: US/10/301.661A  
; PRIOR FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US/09/319,588C  
; PRIOR FILING DATE: 1999-08-27  
; PRIOR APPLICATION NUMBER: FR96/15087  
; PRIOR FILING DATE: 1996-12-09  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1014  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-10-301-661A-6  
Alignment Scores:  
Pred. No.: 9.3e-147 Length: 1014  
Score: 3183.50 Matches: 583  
Percent Similarity: 90.25% Conservative: 74  
Best Local Similarity: 80.08% Mismatches: 58  
Query Match: 69.62% Indels: 13  
Gaps: 4  
US-09-610-313B-31 (1-2463) x US-10-301-661A-6 (1-1014)  
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCGAGCCCGCGAGTTCCCGCAGCGAGCAG 279  
Db 1 PhePheArgGluGluLeuValSerLeuGluThrArgLysLeuProProAspAsn 20  
Qy 280 AAC-----CGCGCAACAGCCCGCAGCGCGCGAGCTGCAGGTG----- 318  
Db 21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40  
Qy 319 ---CGCGGCGAACACCCCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366  
Db 41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60  
Qy 367 CTGAACCTTCCCGCAGATCACCTGTGCGAGCGCCCTGTGAGCATCAAGTGGCGCGC 426  
Db 61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80



; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  
; FILE REFERENCE: 20747Y  
; CURRENT APPLICATION NUMBER: US/09/952,060  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: PCT/US01/28861  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/317,814  
; PRIOR FILING DATE: 2001-09-07  
; PRIOR APPLICATION NUMBER: 60/279,056  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/233,180  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 1350  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Codon optimized gag-1A pol fusion  
US-09-952-060-35

Alignment Scores:  
Pred. No.: 1,22e-143 Length: 1350  
Score: 3119.50 Matches: 599  
Percent Similarity: 81.16% Conservative: 43  
Best Local Similarity: 75.73% Mismatches: 59  
Query Match: 68.22% Indels: 90  
DB: 10 Gaps: 6

US-09-610-313B-31 (1-2463) x US-09-952-060-35 (1-1350)

```
QY 13 ATGCGCGAGCCATGACGAGCCACC---AGCGCAACATCTGTATGACGCGCAGCAAC 69
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 363 LeuAlaGluMetSerGlnValThrAsnSerAlaThrIleMetGlnArgGlyAsn 382
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 70 TTCAAGGGCCCCAAGCGGCATCATCAAGTGTCTCAACTGCGCGCAAGGAGGCGCCACATCGCC 129
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysValGlyHisIleAla 402
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 130 CGCAACTGCGCGCGCCCCCGACAGAGGGTGTCTGGAAGTGGCGGCAAGAGGGCCACCAG 189
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 190 ATCAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCCTTCCCCCAG 249
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 423 MetLysAspCysAsnGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProSerHisLys 442
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 250 GGCAAGGCGCGGAGTTCCCGAGCGAGCAACCGCGCCCAACAGCCCCCAGCGCGAG 309
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 442 sGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProGluGluSe 462
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 310 CTCGAGTGGCGG-----CGGACACCCCCCGAGCGAGCGCGCGCGCGCGAGGCG 363
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 462 rPheArgPheGlyGluGluLysThrThrProSerGlnLysGlnGluProIleAspLys-- 481
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 364 ACCCTGAACCTTCCCCCAGATCACCTCTGGCAGCGCGCCCTGTGTGAGCATCAAGTGGGC 423
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 482 GluLeu-----TyroProLeuAlaSerLeuArg----- 490
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 424 GGCCAGATCAAGGAGGCGCTGCTGGACACCGCGCGCCGACGACACCGTGTGGAGGAGATG 483
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 490 ----- 490
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 484 AGCCTGCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGCGCTTCATCAAGGTG 543
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 490 ----- 490
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 544 CGCCAGTACGACCATCTGTATCGAGATCTGCGGCGAGAGGCCATCGGCACCGTGTCTG 603
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 490 ----- 490
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
QY 604 ATCGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGCACC 663
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 491 -----SerLeuPheGlyAsnAspProSerSerGlnMetAla----- 502
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 664 CTGNACTTCCCATCGAGCCCGGAGAGCGTCCCGTGAAGCTGAAGCCCGGAGCCGAGAC 723
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 503 -----ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAsp 519
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 724 GGCCCAAGGTGAAGCAGTGGCGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCCATC 783
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 520 GlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIle 539
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 784 TGCAGAGATGAGAGAGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTCAAC 843
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 540 CysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyAsn 559
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 844 ACCCCCGTGTTCGCATCAAGAGAGAGGAGCAGCAGCAGGAGTGGCGCAGCTGGTGGACTTC 903
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 560 ThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPhe 579
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 904 CGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGCGCATCCCCACCCC 963
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 580 ArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPro 599
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 964 GCGGCTGAAGAGAGAGAGCGTGACCGTGTGACCGTGGCGAGCGCTACTTCAGC 1023
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 600 AlaGlyLeuLysLysLysSerValThrValLeuAlaValGlyAspAlaTyPheSer 619
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1024 GTGCCCTGAGCAGGAGACTTCCCGAAGTACACCGCTTCCACCATCCCGCAGCATCAAC 1083
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 620 ValProLeuAspGluAspPheArgLysTyThrAlaPheThrIleProSerIleAsnAsn 639
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1084 GAGACCCCGGCGCTCCGCTACCAAGTGTGCTGCCCGAGGCTGGAGGGGAGCGCCC 1143
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 640 GluThrProGlyIleArgTyGlnTyAsnValLeuProGlnGlyTrpLysGlySerPro 659
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1144 AGCATCTTCAGAGCAGCATGACCAAGATCCTCGAGCCCTTCCGCGCCCGCAACCCCGAG 1203
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 660 AlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAsp 679
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1204 ATCGTGTACTACAG-----GCCCGCTGTAGTGGGCGAGCAGCTGGAGATCGGCAG 1257
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 680 IleValIleTyGlnTyMetAlaAlaLeuTyValGlySerAspLeuGluIleGlyGln 699
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1258 CACGCGCCCAAGATCCAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCCACCCGCC 1317
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 700 HisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrPro 719
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1318 GACAAGACCAAGAGGAGCGCCCTTCTGTGGATGGGCTACGAGCTGCAACCCGAC 1377
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 720 AspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyGluLeuHisProAsp 739
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1378 AGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGAGCTGGACCGTGAACGATC 1437
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 740 LysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIle 759
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1438 CAGAAGCTGTGGGCAAGCTGAACCTGGCGCCAGCAGATCTACCCCGGCATCAAGGTGGC 1497
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 760 GlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyProGlyIleLysValArg 779
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1498 CAGCTGTGCAAGTGTGCGCGCGCCAGCCCTGACCCATCGAGCTGGCGCGCTGCAGCGGTG 1557
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 780 GlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGlu 799
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1558 GAGCGCGAGCTGAGCTGCGCGAGAACCGGAGATCTCTCGCGAGCGCCGTGCAGCGGTG 1617
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 800 GluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVal 819
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1618 TACTACGACCCAGCAGGAGACCTGTGGCGCGAGATCCAGAGGAGGCGCCACGACCTGG 1677
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 820 TyrTyAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrp 839
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

QY 1678 ACCTACCAGATCTACAGGAGCCCTTCAAGAACCTGAGACGGCAAGTACGCCAAGATG 1737  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 840 ThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMec 859  
 QY 1738 CGCACCGCCACACCAACGAGTGAAGACGCTGACCGAGGCCGTGCAGAGATCGCCATG 1797  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 860 ArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr 879  
 QY 1798 GAGAGATCGTGATCTGGGGCAAGACCCCAAGTTCGGCTGCCATCCAGAGAGAC 1857  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 880 GluSerIleValIleTyrGlyLysThrProLysPheLysLeuProIleGlnLysGluThr 899  
 QY 1858 TGGGAGACCTGGTGGACCGACTACTCGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTG 1917  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 900 TrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVal 919  
 QY 1918 AACACCCCCCTCGTGAAGCTGTGTACAGCTGGAGAAGGAGGCCCATCATCGCGCGC 1977  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 920 AsnThrProProLeuValLysLeuTyrTyrGlnLeuGluLysGluProIleValGlyAla 939  
 QY 1978 GAGACCTTCTACGTGACCGCGCGCCCAACCCGCGAGACCAAGATCGGCAAGCGCGCTAC 2037  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 940 GluThrPheTyrValAlaGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyr 959  
 QY 2038 GTGACCGACCGGGCGCGAGAGATCGTGAGCTGACCGGAGACCAACCAAGAGAGAC 2097  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 960 ValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThr 979  
 QY 2098 GAGCTGAGGCCATCCAGCTGGCCCTGCGAGGACGCGGCGAGCGGTGAACATCGTGACC 2157  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 980 AlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThr 999  
 QY 2158 GACAGCCAGTACCCCTGGGCGATCATCCAGGCCCGCCGACCAAGAGCGAGCGAGCTG 2217  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1000 AlaSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeu 1019  
 QY 2218 GTCAACCGATCATCGAGCAGCTGTCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCC 2277  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1020 ValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaIleProValPro 1039  
 QY 2278 GCCCACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGGAGAGGGCATCCGC 2337  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1040 AlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArg 1059  
 QY 2338 AAGTGCTGTCTCTGGACGGCATCGAT 2364  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1060 LysValLeuPheLeuAspGlyIleAsp 1068

Search completed: June 2, 2005, 05:28:17  
 Job time : 228.167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 11:33:41 ; Search time 167.424 Seconds  
(without alignments)

11351.628 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 4555

Sequence: 1 gtcagccaccatgcccga.....gggctagcaccggtgaattc 2457

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Deiop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09610313/runat\_31052005\_155136\_15139/app\_query.fasta\_1.7893  
-DB=A-Geneseq\_16Dec04 -QPMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US09610313 @CGN 1 1 557/runat\_31052005\_155136\_15139 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3723.5	81.7	1435	ADS19486	Ads19486 Modified
2	3701.5	81.3	3183	ADP84803	Adp84803 HIV-1 hyb
3	3668.5	80.5	1457	ADN36406	Adn36406 Human pro
4	3577	78.5	1003	AAB69289	Aab69289 HIV-1 non
5	3568	78.3	999	AAB69286	Aab69286 HIV-1 non
6	3563	78.2	1005	AAB69287	Aab69287 HIV-1 non
7	3527	77.4	998	AAE37601	Aae37601 HIV-1 sub
8	3520	77.3	3025	AAB86169	Aab86169 HIV-1 sub
9	3478	76.4	854	AAM48949	Aam48949 HIV-1 sub
10	3474.5	76.3	1000	AAB69282	Aab69282 HIV-1 non

11	3456.5	75.9	1002	2	AAW72993	Aaw72993 HIV isola
12	3454	75.8	1003	6	AAO30963	Aao30963 HIV pol p
13	3454	75.8	1003	8	ADN36413	Adn36413 HIV prote
14	3451.5	75.8	1002	1	AAP81861	Aap81861 Sequence
15	3451	75.8	1003	6	ABR55489	Abr55489 Amino aci
16	3449	75.7	1003	1	AAP60420	Aap60420 Sequence
17	3449	75.7	1003	1	AAP70861	Aap70861 Sequence
18	3443	75.6	1003	3	AAI70602	Aai70602 Codon opt
19	3443	75.6	1003	3	AAI70601	Aai70601 Corrected
20	3441	75.5	1003	5	AAO19387	Aao19387 Lymphaden
21	3435.5	75.4	1003	1	AAR08060	Aar08060 HIV-1 pol
22	3432	75.3	1015	1	AAP60347	Aap60347 HTLV-III
23	3432	75.3	1015	2	AAR43867	Aar43867 HTLV-III
24	3432	75.3	1015	4	AAB85993	Aab85993 Amino aci
25	3429	75.3	1003	3	AAB10047	Aab10047 HIV-1 pol
26	3429	75.3	1003	3	AAI70600	Aai70600 Wild type
27	3428	75.3	1016	2	AAR08062	Aar08062 ACNEV-HIV
28	3425	75.2	1015	2	AAR43875	Aar43875 HTLV-III
29	3421	75.1	1016	2	AAR08063	Aar08063 HIV-1 pol
30	3418	75.0	2033	2	AAR08056	Aar08056 HIV-1 pol
31	3414	75.0	1016	2	AAR08054	Aar08054 HIV-1 pol
32	3412	74.9	1003	3	AAB69284	Aab69284 HIV-1 non
33	3411	74.9	2033	2	AAR08055	Aar08055 HIV-1 pol
34	3410	74.9	1003	8	ADP20076	Adp20076 Human imm
35	3406	74.8	739	5	AAU11874	Aau11874 HIV pol p
36	3406	74.8	1003	1	AAP61508	Aap61508 Sequence
37	3406	74.8	1003	2	AAR29705	Aar29705 pol gene
38	3406	74.8	1003	3	AAI77301	Aai77301 HIV-1 (AT
39	3406	74.8	1003	5	AAE35789	Aae35789 ARV-2 (9B
40	3406	74.8	1003	6	ABU63185	Abu63185 Pol prote
41	3406	74.8	1491	1	AAP91048	Aap91048 Transcript
42	3403	74.7	1003	3	AAB69279	Aab69279 HIV-1 non
43	3401	74.7	1003	6	ABU57552	Abu57552 AIDS abso
44	3399	74.6	1012	6	ABU63325	Abu63325 Human lym
45	3396.5	74.6	1003	2	AAR08059	Aar08059 HIV-1 pol

#### ALIGNMENTS

RESULT 1

ADS19486  
ID ADS19486 standard; protein; 1435 AA.

XX AC ADS19486;  
XX DT 18-NOV-2004 (first entry)  
XX DE Modified HIV-1 GagPol protein.

XX KW Gene delivery; epithelial cell; respiratory system; mammal;  
KW lentiviral packaging system; expression vector; gagpol; gene expression;  
KW lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;  
KW cystic fibrosis; CF; deltaPD; transepithelial potential difference;  
KW gene therapy; HIV-1; mutant; mutein.

XX OS Human immunodeficiency virus 1.  
XX OS Synthetic.

XX PN US2004037780-A1.  
XX PD 26-FEB-2004.  
XX PP 23-AUG-2002; 2002US-00226638.

XX PR 19-NOV-2001; 2001AU-00008942.  
XX PA (PARS/) PARSONS D.  
PA (ANSO/) ANSON D.  
PA (LIMB/) LIMBERIS M.  
PA (FULL/) FULLER M.  
XX PI Parsons D, Anson D, Limberis M, Fuller M;  
XX







QY 13 ATGGCCGAGGCCATGCGCAGCCACCAGC---GCCAATCCTGATCGAGCGCAGCAAC 69  
DB 363 LeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLysGlyAsn 382  
QY 70 TTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGCCACATCGCC 129  
DB 383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402  
QY 130 CGCAATCGCGCGCCCGCCGCAAGAGGCTGCTGGAAGTGCGGCAAGAGGGCCACCAAG 189  
DB 403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422  
QY 190 ATGAGGAGTGCACCGAGCGCCAGCCCAAC----- 219  
DB 423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 442  
QY 220 -----TTCCTCCGCGAGGAC 234  
DB 443 GlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPhePheArgGluAsp 462  
QY 235 CTGGCCTTCCCGAGGCAAGCCCGCGAGTTCGCCAGCGAGCAACCGCGCCCAACAGC 294  
DB 463 LeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSer 482  
QY 294 ----- 294  
DB 483 ProProGluGluSerPheArgPheGlyGluGluThrThrProSerGlnLysGlnGlu 502  
QY 295 -----CCACACGCGCGAGCTGCAGTG----- 318  
DB 503 ProIleAspLysGluLeuTyPProLeuAlaProThrArgArgGluLeuGlnValTrpGly 522  
QY 319 CGCGCGCAACCCCGCAGCAGCGCGCGCGCGAGCCGAGCCAGCGCACCTGAAAC----- 372  
DB 523 ArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSer 542  
QY 373 -----TTCGCCAGATCATCCCTGTGG 393  
DB 543 SerLeuArgSerLeuPheGlySerAspProSerSerGlnPheProGlnIleThrLeuTrp 562  
QY 394 CAGCGCCCTCGTGCAGCATCAAGGTGGCGCGCCAGATCAAGAGGCGCTGTGACACACC 453  
DB 563 GlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThr 582  
QY 454 GCGCGCGACGACACCGTGTGGAGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATG 513  
DB 583 GlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysProLysMet 602  
QY 514 ATCGCGCGCATCGCGGCTTCATCAGGTGCGCCAGTACGACCATCTGATCGAGATC 573  
DB 603 IleGlyGlyIleGlyPheIleLysValArgGlnTyPAspGlnIleLeuIleGluIle 622  
QY 574 TCGCGCAAGAAGCCCATCGGACCGCTGTGATCGCGCCCGCCAGCCCGTGAACATCATCGCC 633  
DB 623 CysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGly 642  
QY 634 CGCAATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCATCAGCCCCCATCGAGACC 693  
DB 643 ArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThr 662  
QY 694 GTGCCGCTGAGCTGAGCCCGGATGACCGGCCCGCCAGAGTGAAGCAGTGGCCCTGACC 753  
DB 663 ValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThr 682  
QY 754 GAGGAGAAGATCAAGCGCCCTGACCGCCATCTCGAGGAGATCGAGAGGAGGGCAAGATC 813  
DB 683 GluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIle 702  
QY 814 ACCAAGATCGCGCCGAGAACCCCTTACAACACCCCGCTTCCGCCATCAAGAGAAGGAC 873  
DB 703 SerLysIleGlyProGluAsnProTyPAsnThrProValPheAlaIleLysLysLysAsp 722

QY 874 AGCAACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGCGCAACCGAGACTTC 933  
DB 723 SerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPhe 742  
QY 934 TGGGAGTGTGAGCTGGGCATCCCCACCCCGCGCTTGAAGAAGAAGAAGACCGTGAACC 993  
DB 743 TrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysSerValThr 762  
QY 994 GTGCTGAGCTGGGGGAGCGCTTACTTACGCTGCCCTGGAGCAGGAGCTTCGCGAAGTAC 1053  
DB 763 ValLeuAspValGlyAspAlaTyPAspValProLeuAspLysAspPheArgLysTyP 782  
QY 1054 ACCGCTTCCACATCCCGCAGCATCAACAACAGACCCCGCGCATCCGCTACCAGTACAAC 1113  
DB 783 ThrAlaPheThrIleProSerIleAsnAsnGlnThrProGlyIleArgTyPTrpGlnTyPAsn 802  
QY 1114 GTGCTGCCCGCAGGGGTGAAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATC 1173  
DB 803 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThrLysIle 822  
QY 1174 CTGAGAGCCTTCCGGGCGCGCAACCCCGAGATCGTGATCTACCAG-----GCCCGCTG 1227  
DB 823 LeuGluProPheArgLysGlnAsnProAspIleValIleTyPTrpGlnTyPMetAspAspLeu 842  
QY 1228 TACGTGGGCGAGCAGCTGAGATCGCGCAGCACCGCGCAAGATCGAGAGCTGCGCAAG 1287  
DB 843 TyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGln 862  
QY 1288 CACTGTGCTGCGCTGGGGCTTCCACCAACCCCGCAGCAAGAGAGCAGCAAGAGGAGCCCGCTTC 1347  
DB 863 HisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPhe 882  
QY 1348 CTG-----CCATCGAGCTGCACCCCGACAAAGTGGACCGTGCACCCATCGAGCTGCC 1401  
DB 883 LeuTrpMetGlyTyPTrpGluLeuHisProAspLysTrpThrValGlnProIleValLeuPro 902  
QY 1402 GAGAGGAGAGCTGAGCCGTGAACGACATCAGAGCTGTGGGCAAGCTGAACCTGGGCGC 1461  
DB 903 GluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAla 922  
QY 1462 AGCCAGATCTTACCCCGCATCAAGTGGCGAGCTGTGCAAGCTGCTCGCGGCGCCCAAG 1521  
DB 923 SerGlnIleTyPAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLys 942  
QY 1522 GCGCTGACCGACATCGTCCCTGACCGAGAGCGCGAGCTGAGCTGGCCGAGAACCGC 1581  
DB 943 AlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArg 962  
QY 1582 GAGATCTCGCGAGCGCGTGCACGGCTGTACTACGACCCCGCAGCAAGACCTGTGTGGC 1641  
DB 963 GluIleLeuLysGluProValHisGlyValTyPAspProSerLysAspLeuIleAla 982  
QY 1642 GAGATCCAGAAGCAGGCGCCACGACGAGTGGACCTTACAGATCTTACAGAGCCCTTCAAG 1701  
DB 983 GluIleGlnLysGlnGlyGlnTyPTrpThrTyPTrpGlnIleTyPTrpGlnProPheLys 1002  
QY 1702 AACCTGAAGACCGCAAGTACGCCAAGATGGCAGCGCCCGCCACCAACGACGCTGAAGCAG 1761  
DB 1003 AsnLeuLysThrGlyLysTyPAlaArgMetLysGlyAlaHisThrAsnAspValLysGln 1022  
QY 1762 CTGACCGAGGCGGTGCAGAAGATCGCCATCGCAGAGCATCGTGTGCGGCAAGACCCCGC 1821  
DB 1023 LeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPro 1042  
QY 1822 AAGTTCGCGCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAG 1881  
DB 1043 LysPheLysLeuProIleGlnLysGluThrTrpGluAlaTyPTrpThrGluTyPTrpGln 1062  
QY 1882 GCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGGGTAC 1941  
DB 1063 AlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyP 1082  
QY 1942 CAGCTGGAGAAGGAGCCCATCATCGCGCGCAGACCTTCTACGTGGACGCGCGCCCAAC 2001

```

Db 1083 GlnLeuGluYsGluProIleGlyAlaGluThrPheYrValAspGlyAlaAlaAsn 1102
QY 2002 CGCAGACCAAGATCGGCAAGCGCGGTACGTGACCGACCGCGCGCGCAAGATCGTG 2061
Db 1103 ArgGluThrYsLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 1122
QY 2062 AGCTGACCGAGACCAACCAAGACCGAGCTGCGAGGCCATCCAGCTGGCCCTGCAG 2121
Db 1123 ProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142
QY 2122 GACAGCGCGCAGGAGTGAACATCGTACCGACGACGACGATCGAGCTGCGATCCAG 2181
Db 1143 AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleGln 1162
QY 2182 GCCCAGCGCGACAAAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCTGATCAAG 2241
Db 1163 AlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuLys 1182
QY 2242 AAGGAGAGGTGTACCTGAGCTGGTGGTCCCGCCCAAGGGCATCGCGCGCAAGCAGCAG 2301
Db 1183 LysGluLysValTyrLeuAlaTyrValProAlaHisLysGlyIleGlyGlyAsnGluGln 1202
QY 2302 ATCGACAGCTGTGAGCAGGCGCATCGCGAAGTGTCTTCTCGAGCGGATCGAT 2358
Db 1203 ValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1221

RESULT 3
ADN36406
ID ADN36406 standard; protein; 1457 AA.
AC
XX ADN36406;
DT 15-JUL-2004 (first entry)
DE Human protein for anti-HIV vaccine.
XX
XX anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
XX Homo sapiens.
XX
XX WO2004035006-A2.
XX
XX 29-APR-2004.
XX
XX 17-OCT-2003; 2003WO-US033112.
XX
XX 18-OCT-2002; 2002US-0419465P.
XX
XX (AARO-) AARON DIAMOND AIDS RES CENT.
XX
XX Huang Y, Ho DD, Chen Z;
XX
XX WPI; 2004-348328/32.
XX
XX N-PSDB; ADN36405.
XX
XX Nucleic acid vector comprising at least one HIV sequence operably linked
XX to a promoter and encoding a protein that does not assemble into viral
XX particles, useful in immunizing a subject against HIV infection.
XX
XX Disclosure; SEQ ID NO 20; 166pp; English.
XX
XX The invention relates to a nucleic acid vector comprising at least one
XX HIV sequence operably linked to a promoter and encoding a protein that
XX does not assemble into viral particles. The nucleic acid vector is useful
XX in immunizing a subject against HIV infection. This sequence corresponds
XX to a peptide used in the invention.
XX
XX SQ Sequence 1457 AA;

```

Alignment Scores:

Pred. No.: 1e-203 Length: 1457  
Score: 3668.50 Matches: 702

```

Percent Similarity: 91.36% Conservative: 28
Best Local Similarity: 87.86% Mismatches: 47
Query Match: 8 Indels: 23
DB: 5 Gaps: 5
US-09-610-313B-32 (1-2457) x ADN36406 (1-1457)
QY 13 ATGCGCGAGCCATGATGACCGGCGCCAGCCCAACATCTGTATGCGAGCGCAACTTC 72
Db 383 LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnPhe 401
QY 73 AAGGCGCCCAAGCGCATCATCAAGTCTTCACTGCGCGCAAGGAGGCGCCATCGCCCGC 132
Db 402 LysGlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 421
QY 133 AACTGCGCGCCCGCCCAAGAGAGGGCTGCTGGAGGTGCGCAAGGAGGGGCCACAGATG 192
Db 422 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 441
QY 193 AAGGACTGTCACCGAGCGCGCCAGGCCAACTTCTTCGCGAGGACCTTCCCCCGAGGCG 252
Db 442 LysAspCysThrGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProProHisLysG 461
QY 253 AAGCGCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCCCAACAGCCCGCGAGCGAGTG 312
Db 461 YArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProAlaGluSerPh 481
QY 313 CAGGTGCGCGCGACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db 481 eGlyPheGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuTh 501
QY 373 TTC-----CCCCAGATCACCCCTGTGGCA 395
Db 501 rSerLeuLysSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpG 521
QY 396 GCGCGCCCTGTGATGATCAAGTGGCGCGCGAGATCAAGAGGCGCCCTGCTGGACACCGG 455
Db 521 nArgProLeuValSerIleArgValGlyGlyGlnIleLysGluAlaLeu----- 538
QY 456 CGCGCAGCAGCACCGTCTGCGAGGAGATGAGCTGCGCGCAAGTGGAGCCCAAGATGAT 515
Db 539 ----AspAspThrValLeuGluValAsnLeuProGlyLysTrpLysProLysMecIl 557
QY 516 CGCGCGCATCGCGCGCTTCATCAAGGTGCGCGAGTACGACAGATCTGTGATCGAGATCG 575
Db 557 eGlyGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleProIleGluIleCy 577
QY 576 CGCGAAGAGCCATCGGCAACCGTCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 635
Db 577 sGlyLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597
QY 636 CAACATGCTGACCGCGCTGCGCGCTGACCCCTGACCTTCCCATCAGCCCGCGCGCGCG 695
Db 597 gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrI 617
QY 696 GCCCGTGAAGTGAAGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755
Db 617 eProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrG 637
QY 756 GGAGAAGATCAAGCGCTGACCGCGCATCTGCGAGGAGATGGAGAGGAGGAGGAGATCAC 815
Db 637 uGluLysIleLysAlaLeuThrAlaIleCysAspGluMetGluLysGluGlyLysIleTh 657
QY 816 CAGATCGCGCGCGCGAGACCCCTACACACCGCGCGCTGCTGCGCATCAAGAGAGAGGAG 875
Db 657 rLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSe 677
QY 876 CACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAAACAGCGCGCGCGCGCGCGCT 935
Db 677 rThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTr 697
QY 936 GGAGGTGACGTGGGCATCCCCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 995

```

Db 697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrVa 717  
QY 996 GCTGACGTGGCGAGCGCTACTTTCAGCGTGCCTTGGACGAGGACTTCCGCAATACAC 1055  
Db 717 IleuAspValGlyAspAlaTyPheSerValProLeuTyrGluAspPheArgLysTyThr 737  
QY 1056 GCCTTTACCATCCCGAGCATCAACACGAGACCCCGGCATCCGCTACCGTACAAGCT 1115  
Db 737 rAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyAsnVa 757  
QY 1116 GCTGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175  
Db 757 IleuProGlnGlyTyPlysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777  
QY 1176 GSAGCCCTTCCGCGCGCGCAACCCGAGATCGTGATCTACGAG-----GCCCCCTGTA 1229  
Db 777 uGluProPheArgAlaGlnAsnProGluIleValIleTyrGlnTyGlyAspAspLeuTy 797  
QY 1230 CGTGGCGAGCGACCTGGAGATCGGCAGCACCGCCGCAAGATCGAGGAGCTGCGCAAGCA 1289  
Db 797 rValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuArgGluHI 817  
QY 1290 CTGCTGGCGTGGGCTTTCACACCCCGCAGCAAGAACCAAGAGGAGCCCGCTTCTCT 1349  
Db 817 sLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLe 837  
QY 1350 G-----CCCATCGAGCTGCACCCCGACAGTGGACCGTGCAGCCCATCGAGCTGCCGA 1403  
Db 837 TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGI 857  
QY 1404 GAAGGAGAGCTGACCGTGAACGACATCCAGAGCTGGTGGCAAGCTGAACCTGGCGCAG 1463  
Db 857 uLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877  
QY 1464 CCAGATCTACCCCGGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGCGCGCAAGGC 1523  
Db 877 rGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897  
QY 1524 CTTGACCGACATCGTGCCTGACCGAGAGCGCGCGAGCTGGAGCTGGCGGAGAACCGGA 1583  
Db 897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGI 917  
QY 1584 GATCTCGCGAGCGCGTGCAGCGGTGTACTACGACCCCGACAGGACCTGGTGGCGCGA 1643  
Db 917 uLeuLeuLysGluProValHisGlyAlaTyTrpAspProSerLysAspLeuIleAlaGI 937  
QY 1644 GATCAGAGCAGGCGCCACGACCTGACGACCTTACAGATCTACAGAGGCGCTTCAAGAA 1703  
Db 937 uIleGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAs 957  
QY 1704 CTTGAAGACCGCAAGTACCGCAAGATGGCGACCGCCCGCACACCAACGACGTGAAGCAGCT 1763  
Db 957 nLeuLysThrGlyLysTyAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977  
QY 1764 GACCGAGCGCGTGCAGAGATCGCATCGAGAGCATCGTGATCTCGGCGAAGACCCCAA 1823  
Db 977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTrpGlyLysIleProHy 997  
QY 1824 GTTCGCTGCTCCATCCAGAGGAGACCTGGGAGACCTGGTGAGCCGACTACTGGCAGGC 1883  
Db 997 sPheArgLeuProIleProLysGluThrTrpGluThrArgTrpThrAlaTyTrpGlnAl 1017  
QY 1884 CACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGTGTACA 1943  
Db 1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyTrGI 1037  
QY 1944 GCTGAGAGGAGCCCATCATCGCGCGGAGACCTTCTAGTGGACGCGCGCCCAACCG 2003  
Db 1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTyValAspGlyAlaAlaAsnAr 1057  
QY 2004 CGAGACCAAGATCGGCAAGGCGGCTAGCTGACCGACCGCGCGCGCAGAGATCGTAG 2063  
Db 1057 gGluThrLysMetGlyLysAlaGlyTyValThrAspArgGlyArgGlnLysIleValSe 1077

QY 2064 CCTGACGAGACCCACCAACGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTCAGGA 2123  
Db 1077 rLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleCysLeuAlaLeuGlnAs 1097  
QY 2124 CAGCGGACGAGGTGAACATCGTACCGACGACGACGACGACGACGACGACGACGACGACG 2183  
Db 1097 pSerGlySerGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIleIleGlnAl 1117  
QY 2184 CCAGCCCGACAGAGCGAGCGAGCTGGTGAACAGATCATTCAGAGCAGCTGATCAAGAA 2243  
Db 1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuLysIle 1137  
QY 2244 GGAGAGGTGTACTGAGCTGGTGGTCCCGCCCAAGGGCATCGCGGCAACGACGACAT 2303  
Db 1137 sGluArgValTyrLeuSerTrpValProAlaHisGlyIleGlyGlyAsnGluGlnVa 1157  
QY 2304 CGAAGAGCTGTGAGCAAGGCGCATCCGCAAGGTGTCTCTGACGCGCATCGAT 2358  
Db 1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175

RESULT 4  
AAB69289  
ID AAB69289 standard; protein; 1003 AA.  
XX AAB69289;  
XX 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX HIV-1 non-subtype B clone 94IN476-104 pol protein.  
XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
KW vif; vpr; tat; rev; nef; vaccine.  
XX Human immunodeficiency virus 1.  
XX WO200026416-A1.  
XX 11-MAY-2000.  
XX 25-OCT-1999; 99WO-US024837.  
XX 02-NOV-1998; 98US-00184418.  
XX (UABR-) UAB RES FOUND.  
XX Hahn BH, Shaw GM, Gao F;  
XX WPI; 2000-365651/31.  
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.  
XX Claim 41; Fig 15; 131pp; English.  
XX The present invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX SQ Sequence 1003 AA;  
Alignment Scores:  
Pred. No.: 1,8e-198 Length: 1003  
Score: 3577.00 Matches: 673  
Percent Similarity: 97.49% Conservative: 26  
Best Local Similarity: 93.86% Mismatches: 14

Query Match: 78.53% Indels: 4  
 DB: 3 Gaps: 2  
 US-09-610-313B-32 (1-2457) x AAB69289 (1-1003)

220 TTCTTCGCGAGGACCTTCCCGCCAGGCGCGGAGTTCCTCCCGAGCGAG 279  
 Db 1 PhePheArgGluAenLeuAlaPheProGlnGlyGluAlaArgGluPheProSerLysGln 20  
 280 AACCGCGCAACAGCCCCACAGCGCGGAGCTGAGGTGCGCGCGCAACACCCCGGAGC 339  
 Db 21 AlaArgAlaAenSerProThrSerArgGluLeuGlnValGlnGlyAAsnProArgSer 40  
 340 GAGCGCGCGCGAGCGCGAGCGGACCTGAACTTCCCGCCAGATCACCTGTGGAGCGC 399  
 Db 41 GluAlaGlyValGluArgGlnGlyThrLeuAenPheProGlnIleThrLeuTrpGlnArg 60  
 400 CCCCTGTGTGACATCAGGTGGCGGCGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATCGGC 519  
 Db 81 AspAspThrValLeuGluGluLeuAlaLeuProGlyArgTrpLysProLysMetIleGly 100  
 520 GGCAATCGCGCGCTTCATCAAGGTGGCGGAGTACGACCAAGATCTCTGATCGAGATCTGCGGC 579  
 Db 101 GlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGluLeuCysGly 120  
 580 AAGAAGCGCATCGGACCGCTGCTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 639  
 Db 121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAenIleIleGlyArgAsp 140  
 640 ATGTGTGACCGAGTGGGTGACCTGAACTTCCCATCTCCAGCCCATCGAGACCGTGC 699  
 Db 141 MetLeuThrGlnLeuGlyCysThrLeuAenPheProIleSerProIleGluThrValPro 160  
 700 GTGAAGTGAAGCGCGGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 759  
 Db 161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180  
 760 AAGATCAAGGCGCTGACCGCCATCTGCGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAG 819  
 Db 181 LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys 200  
 820 ATCGGCGCGGAGAACCTTACAACACCCCGTGTTCGCCATCAAGAGAGAGAGAGAGAGAG 879  
 Db 201 IleGlyProGluAenProTyraThrProValPheAlaIleLysArgLysAspSerThr 220  
 880 AAGTGGCGCAAGCTGTGGACTTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAG 939  
 Db 221 LysTrpArgLysLeuValAspPheArgGluLeuAenLysArgThrGlnAspPheTrpGlu 240  
 940 GTGAGCTGGCGATCCCGCCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 999  
 Db 241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 260  
 1000 GAGTGGCGGACGCTACTTACAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1059  
 Db 261 AspValGlyAspAlaTyPheSerValProLeuAspGluGlyPheGlyLysTyThrAla 280  
 1060 TTCAACATCCCGAGCATCAACAGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1119  
 Db 281 PheThrIleProSerIleAenAenGluThrProGlyIleArgTyGlnTrpAenValLeu 300  
 1120 CCCAGGCGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1179  
 Db 301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320  
 1180 CCCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1233  
 Db 321 PropheArgAlaArgAenProLysIleValIleTyThrGlnTyMetAspAspLeuTyVal 340

1234 GGCAGCGACCTGAGAGATCGCGCAGACCGCGCCCAAGATCGAGGAGCTCGCGAAGCACCTG 1293  
 Db 341 GlySerAspLeuGluIleGlyHisArgAlaLysIleGluLeuArgAlaHisLeu 360  
 1294 CTGCGCTGGGCTTCCACACCCCGCAGCAAGAGCACCAAGAGGAGGCGGCGGCGGCGGCGG 1350  
 Db 361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380  
 1351 ---CCCATCGAGCTGCACCCCGCAGCAAGTGAACCTGCGCGCCCATCGAGCTGCCCGAGAG 1407  
 Db 381 MetGlyTyThrGluLeuHisProAspLysTrpThrValGlnProIleLysLeuProGluLys 400  
 1408 GAGAGCTGAGCCCGTGAACAGCATCCAGAGCTGTGTGGCAAGCTGAACCTGGCGCAGCCAG 1467  
 Db 401 AspSerTrpThrValAenAspIleGlnLysLeuValGlyLysLeuAenTrpAlaSerGln 420  
 1468 ATCTACCCCGGCATCAAGTGGCGGCGGAGCTGCAAGCTGCGCGGCGGCGGCGGCGGCGG 1527  
 Db 421 IleTyProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440  
 1528 ACCGACATCTGCTGCCCTCGACCGAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATC 1587  
 Db 441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAenArgGluLeu 460  
 1588 CTGCGCGAGCGCGTGCACGCGCTGTACTACGACCCCGCAGCAAGAGGAGCTGTGTGGCGGAGATC 1647  
 Db 461 LeuLysGluProValHisGlyValTyThrAspProSerLysAspLeuIleAlaGluLeu 480  
 1648 CAGAGCGGCGGCGGCGGAGTGGACCTTACAGATCTTACAGGAGCGGCGGCGGCGGCGGCGG 1707  
 Db 481 GlnLysGlnGlyHisAspGlnTrpThrTyroGlnIleTyroGlnGluProPheLysAenLeu 500  
 1708 AAGACCGCGCAAGTACCGCAAGATCGCACCGCCCGCACCAACAGCGTGAAGCAGCTGAC 1767  
 Db 501 LysThrGlyLysTyroAlaLysMetArgThrAlaHisThrAenAspValLysGlnLeuThr 520  
 1768 GAGCGCTGCAAGATCGCCATCGAGAGCATCTGTGATCTGGGCGAAGACCCCGCAAGTTC 1827  
 Db 521 GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp\*\*LysThrProLysPhe 540  
 1828 CGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGGCGGCGGCGGCGGCGGCGG 1887  
 Db 541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyroTrpGlnAlaThr 560  
 1888 TGGATCCCGAGTGGGAGTTCGTGAACACCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1947  
 Db 561 TrpIleProAspTrpGluPheValAenThrProProLeuValLysLeuTrpTyroGlnLeu 580  
 1948 GAGAAGAGCGCCATCATCGCGCGGAGACCTTCTACGTGGACGCGCGGCGGCGGCGGCGGCGG 2007  
 Db 581 GluLysGluProIleValGlyAlaGluThrPheTyroValAspGlyAlaAlaAenArgGlu 600  
 2008 ACCAAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2067  
 Db 601 ThrLysValGlyLysAlaGlyTyroValThrAspArgGlyArgGlnLysIleValSerLeu 620  
 2068 ACCGAGACCAACACAGAGAGAGCGAGCTGCAGGCGCATCCAGCTGGCGGCGGCGGCGGCGG 2127  
 Db 621 ThrGluThrThrAenGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 640  
 2128 GGCAGCGAGGTGAACATCTGTGACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2187  
 Db 641 GlyThrGluValAenIleValThrAspSerGlnTyroAlaLeuGlyIleIleGlnAlaGln 660  
 2188 CCGCAGAGAGCGGAGCGGAGCTGGTGAAACAGATCATCGAGGAGCTGTGATCAAGAGAGAG 2247  
 Db 661 ProAspLysSerGluSerGluLeuValAenGlnIleIleGluGlnLeuIleAenLysGlu 680  
 2248 AAGTGTACTGAGCTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2307  
 Db 681 ArgValTyroLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnGluValAsp 700  
 2308 AAGCTGTGAGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2358



Db 701 ArgLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleasp 717  
 :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::

RESULT 5  
 AAB69286  
 ID AAB69286 standard; protein; 999 AA.  
 XX AAB69286;  
 XX AC AAB69286;  
 XX DT 12-SEP-2003 (revised)  
 XX DT 20-APR-2001 (first entry)  
 XX DE HIV-1 non-subtype B clone 96ZM651-8 pol protein.  
 XX KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 XX KW vif; vpr; tat; rev; nef; vaccine.  
 XX OS Human immunodeficiency virus 1.  
 XX PN WO200026416-A1.  
 XX PD 11-MAY-2000.  
 XX PF 25-OCT-1999; 99WO-US024837.  
 XX PR 02-NOV-1998; 98US-00184418.  
 XX PA (UABR-) UAB RES FOUND.  
 XX PI Hahn BH, Shaw GM, Gao F;  
 XX DR WPI; 2000-365651/31.  
 XX PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.  
 XX PS Claim 41; Fig 15; 131pp; English.  
 XX CC The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
 XX SQ Sequence 999 AA;  
 Alignment Scores:  
 Pred. No.: 5,976-198 Length: 999  
 Score: 3568.00 Matches: 673  
 Percent Similarity: 97.21% Conservative: 24  
 Best Local Similarity: 93.86% Mismatches: 16  
 Query Match: 78.33% Indels: 4  
 Db: 3 Gaps: 2  
 US-09-610-313B-32 (1-2457) x AAB69286 (1-999)  
 QY 220 TTCTTCCGGAGGACTGGCTTCCCGGCAAGGCGCGAGTCCCGAGCGAGCGAG 279  
 Db 1 PhePheArgLysValLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20  
 QY 280 AACCGCGCCACAGCCCAACAGCGCGAGCTGCGAGTGCAGGCGCGGCGCAACACCCCGCAGC 339  
 Db 21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSer 40  
 QY 340 GAGGCGGCGCGAGCGCGAGCGCGAGCGCGAGCTGCGAGTCCCGCGAGTCCCGCGAGCGAG 399  
 Db 41 GluAlaGlyValGluArgGlnGlySerLeuAsnPheProGlnIleThrLeuTrpGlnArg 60  
 QY 400 CCCTGGTGAGCATCAAGGTGGCGCGCCAGATCAAGGAGGCGCTGTGTGACACCGCGCGC 459

Db 61 ProLeuValSerIleLysValGlyGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80  
 QY 460 GACGACACCGTCTGGAGGAGATGAGCTGCCCGCAAGTGGAAAGCCCAAGATGATCGGC 519  
 Db 81 GlyAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGly 100  
 QY 520 GGCATCGGGCGCTTCATCAAGGTGGCGGACAGTACGACAGATCCTGATCGAGATCTGCGGC 579  
 Db 101 GlyIleGlyGlyPheIleGluValArgGlnTrpAspGlnIleProMetGluIleCysGly 120  
 QY 580 AAGAGGCGCATCGGCACCGTGTGATCGGCGCCCGCCCGCGTGAACATCATCGCGCGCAAC 639  
 Db 121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 140  
 QY 640 ATGCTGACCGAGCTGGGCTGACCTGAACTTCCCATCAGCCCATCGAGACCGTGC 599  
 Db 141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160  
 QY 700 GTGAGCTGAAGCGCGCATGAGCGCGCCCAAGGTGAAGCAGTGGCGCCCTGACCGGAG 759  
 Db 161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180  
 QY 760 AAGATCAAGGCGCTGACCGCCATCTCGAGGAGATGAGAGAGGCGCAAGATCACCAG 819  
 Db 181 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyIleThrLys 200  
 QY 820 ATCGGCGCGGAGAACCTTACAACACCCCGTGTTCGCATCAAGAGAGAGACGAC 879  
 Db 201 IleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLysAspSerThr 220  
 QY 880 AAGTGGCGGAGCTGGTGGACTTCCCGGAGCTGAACAGCGCACCGAGACTTCTGGGAG 939  
 Db 221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240  
 QY 940 GTGACGTGGGCGATCCCGCACCCCGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 999  
 Db 241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu 260  
 QY 1000 GACGTGGGCGGCGCTTACTTACGTGGCGCTGGAGAGAGCTTCCGCGAGTACCGCC 1059  
 Db 261 AspValGlyAspAlaTyPheSerValProLeuAspGluSerPheArgLysTrpThrAla 280  
 QY 1060 TTCACATCCCGAGCATCAACAGAGACCGCGCATCCGCTACAGTACACAGTGTG 1119  
 Db 281 PheThrIleProSerThrAsnAsnGluThrProGlyIleArgTrpGlnTrpAsnValLeu 300  
 QY 1120 CCCCAGGCGTGAAGGCGAGCGCGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAG 1179  
 Db 301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320  
 QY 1180 CCCTTCCCGCGCGCAACCCGAGATCGTGTATCTACAG-----GCCCGCTGTACGTG 1233  
 Db 321 ProPheArgAlaGlnAsnProAspIleValIleTyGlnTrpMetAspLeuTyVal 340  
 QY 1234 GCGAGCGACCTGGAGATCGGCGAGCGCGCGAGATCGAGGAGCTGCGAGGAGCGCTG 1293  
 Db 341 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeu 360  
 QY 1294 GTGCGCTGGGCTTCCACCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1350  
 Db 361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeuTrp 380  
 QY 1351 ---CCCATCGAGCTGCACCCCGCAAGTGGAGCGCTGCGAGCGCGCTGCGAGAGAG 1407  
 Db 381 MetGlyTyTrpGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuAlaGluLys 400  
 QY 1408 GAGAGCTGAGCGTGAACGAGATCCAGAGCTGGGGCGCAAGCTGAACCTGGCGCGAG 1467  
 Db 401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpLysGln 420  
 QY 1468 ATCTACCGCGCGCATCAAGGTGGCGCGAGCTGTGCAAGAGCTGTGCGCGCGCGCAAG 1527

Db 421 IleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440  
 Qy 1528 ACCGACATCGTGGCCCTCACCAGAGAGCCGAGCTGGAGCTGGCGAGAACCGGAGATC 1587  
 Db 441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnLysGluIle 460  
 Qy 1588 CTGGCGAGCCCGTGCACGGGTGTACTACGACCCAGCAAGACCTGGTGGCGAGATC 1647  
 Db 461 LeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIle 480  
 Qy 1648 CAGAAGCAGGCGCCAGCAGCTGGACCTACCAGATCTACCAAGGAGCCCTTCAAGAACCTG 1707  
 Db 481 GlnLysGlnGlyHisAspGlnTrpTyrGlnIleTyrGlnGluProPheLysAsnLeu 500  
 Qy 1708 AAGACCGCAAGTAGTACGCCAAGATCGCACCCGCCACCAACGACGCTGAAGCAGCTGACC 1767  
 Db 501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520  
 Qy 1768 GAGCGCTGCAGAGATGCCATCGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTC 1827  
 Db 521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTrpGlyLysIleProLysPhe 540  
 Qy 1828 CGCTGCTCCATCCAGAGGAGACCTGGGAGACCTGTGGACCGACTACTGGCAGGCCACC 1887  
 Db 541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAlaThr 560  
 Qy 1888 TGGATCCCGGAGTGGAGTTCGTGAACACCCGCCCTGGTGAAGCTGTGTACAGCTG 1947  
 Db 561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTrpTyrGlnLeu 580  
 Qy 1948 GAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGACGGCGCGCCCAACCGCGAG 2007  
 Db 581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600  
 Qy 2008 ACCAGATCGCAGAGCGCGGTACTGTGACACCGGGCCGCGAGAGATCGTGAGCTG 2067  
 Db 601 ThrLysLeuGlyLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620  
 Qy 2068 ACCGAGACCCACCAACAGAACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGACAGC 2127  
 Db 621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640  
 Qy 2128 GGCAGCAGGTGAACATCGTACCCAGACAGCCAGTACGCGCTTGGCGCATCATCCAGGCCAG 2187  
 Db 641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660  
 Qy 2188 CCCCACAGACGAGAGCGAGCTGGTGAACACGATCATCGAGCAGCTGATCAAGAGGAG 2247  
 Db 661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuLysLysGlu 680  
 Qy 2248 AAGGTGTACCTGAGCTGGTGGTCCCGCCACCAAGGGCATCGCGGCAACGAGCAGATCGAC 2307  
 Db 681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysAsnGluGlnValAsp 700  
 Qy 2308 AAGCTGTGTGACAAAGGGCATCCGCAAGGTGTGTCTTCTGGAGCGCATCGAT 2358  
 Db 701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

## RESULT 6

AA69287  
 ID AA69287 standard; protein; 1005 AA.  
 AC AA69287;  
 XX  
 XX  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 XX  
 XX HIV-1 non-subtype B clone 962M751-3 pol protein.  
 XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpr;  
 KW vif; vpr; tat; rev; net; vaccine.  
 XX  
 XX Human immunodeficiency virus 1.  
 OS

XX WO200026416-A1.  
 FN 11-MAY-2000.  
 PD 25-OCT-1999; 99WO-US024837.  
 XX 02-NOV-1998; 98US-00184418.  
 PR (UABR-) UAB RES FOUND.  
 PA Hahn BH, Shaw GM, Gao F;  
 PI WPI; 2000-365651/31.  
 DR Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.  
 XX Claim 41; Fig 15; 131pp; English.  
 PS The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection.. (Updated on 12-SEP-2003 to standardise OS field)  
 XX SQ Sequence 1005 AA;

## Alignment Scores:

Pred. No.: 1.16e-197 Length: 1005  
 Score: 3563.00 Matches: 677  
 Percent Similarity: 96.13% Conservative: 18  
 Best Local Similarity: 93.64% Mismatches: 18  
 Query Match: 78.22% Indels: 10  
 DB: 3 Gaps: 3

US-09-610-313B-32 (1-2457) x AAB69287 (1-1005)

Qy 220 TTCTTCCGAGGACCTGGCTTCCCGGAGCAAGCCCGGAGTTCCCGAGGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyLeuProSerGluGln 20  
 Qy 280 AACGCGCC-----AACGCGCCACCGAGCCGCGAGCTGCAGGTGCGC 321  
 Db 21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40  
 Qy 322 GCGGACAAACCCCGCAGCGAGCGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCG 381  
 Db 41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60  
 Qy 382 ATCAACCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGCGCGCGAGCGCGAGCG 441  
 Db 61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnLysGluAla 80  
 Qy 442 CTGCTGGACACCGCGCGCGAGCAGCACCGTGTGTGAGGAGATGAGCTGCCCGGCAAGTGG 501  
 Db 81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrp 100  
 Qy 502 AAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGGCGCGAGCGCGAGATC 561  
 Db 101 LysProLysMetIleGlyIleGlyPheLysValArgGlnTyrAspGlnIle 120  
 Qy 562 CTGATCGAGATCTGCGGCAAGAAGCGCATCGGACCGCTGTGTGATCGGCGCGCGAGCG 621  
 Db 121 LeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVal 140  
 Qy 622 AACATCATCGCGCGCGACATGTCGACCGAGTGGGTGCGACCTGTGAATCTTCCCATCAGC 681  
 Db 141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160

QY 682 CCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGGACGGCCCAAGGTGAAGCAG 741  
DB 161 ProIleGluThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGln 180  
QY 742 TGGCCCTGACCGAGGAGATCAAGCCCTGACCGCCATCTGCGAGGAGATGAGAAAG 801  
DB 181 TrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLys 200  
QY 802 GAGGCAAGATCACCAAGATCGGCCCGCCGAGAACCCCTACAACACCCCGTGTTCGCCATC 861  
DB 201 GluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIle 220  
QY 862 AAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGTGGACTTCGCGAGCTGAACAAGCGC 921  
DB 221 LysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgLysLeuLysLysArg 240  
QY 922 ACCCAGGACTTCTGGAGGTGCAGCTGGCGCATCCCGCCACCGCCGCTGAAGAAGAG 981  
DB 241 ThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLys 260  
QY 982 AAGACGTGACCGTCTGTCGAGCTGGCGACGCTACTTCAGCGTCCCGCTGGACGAGGAC 1041  
DB 261 LysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGly 280  
QY 1042 TTCGCCAAGTACACCGCTTACCATCCCGACATCAACACGAGACCCCGGCATCCGC 1101  
DB 281 PheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArg 300  
QY 1102 TACCAAGTACAACGCTGCGCCCGAGGGCTGGAAGGCGAGCCCGCATCTTCAGAGCAGC 1161  
DB 301 TyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSer 320  
QY 1162 ATGACCAAGATCCTGAGCCCTTCGCGCCCGCCCAACCCCGAGATCGTGTATCTACACG--- 1218  
DB 321 MetIleLysIleLeuGluProPheArgThrGlnAsnProGluIleValIleTyrGlnTyr 340  
QY 1219 ---GCCCCCTCTAGTGGGCGAGCACCTGGAGATCGGCGCACACCGCGCCAAAGATCGAG 1275  
DB 341 MetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlu 360  
QY 1276 GAGCTGCGCAGACACCTGTGCGCTGGGGCTTACCACCCCGCACAAGAAGCACGAGAAG 1335  
DB 361 GluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLys 380  
QY 1336 GAGCCCGCTTCTCGT-----CCCATCGAGCTGACCCCGCAAGTGAAGCCGTGAGCCG 1389  
DB 381 GluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPro 400  
QY 1390 ATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAAGCTGTGGGCAAG 1449  
DB 401 IleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLys 420  
QY 1450 CTGAATGGCGCCAGCAGATCTACCCCGCATCAAGGTGCGCGCAGCTGTGCAAGCTGTG 1509  
DB 421 LeuAsnTrp\*\*SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeu 440  
QY 1510 CCGCGCGCCCAAGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCGCGAGCTGGAGCTG 1569  
DB 441 ArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeu 460  
QY 1570 GCGGAGAACCGGAGATCTGCGCGCGCCCGTGCAGCGGCGTGTACTAGCACCCCGCAGAG 1629  
DB 461 AlaGluSerArgGluIleLysLysGluProValHisGlyValTyrTrpAspProSerLys 480  
QY 1630 GACCTGCTGGCGCAGATCCAGAAAGCAGAGGCGCCACGACCACTGAGATCTACCAAG 1689  
DB 481 AspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnValTyrGln 500  
QY 1690 GAGCCCTTCAAGAACCTGAAGACCGGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1749  
DB 501 GluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsn 520

QY 1750 GACGTGAAGACGATGACCGAGCCCGTGCAGAAAGATGCGCATGGAGAGCATCGTGTCTGG 1809  
DB 521 AspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrp 540  
QY 1810 GCGAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAGACCTGGAGACCTGGTGACC 1869  
DB 541 GlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThr 560  
QY 1870 GACTACTGGCAGGCGCACCTGATCCCGAGTGGAGTTCGTGAACACACCCCGCTGGTG 1929  
DB 561 AspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVal 580  
QY 1930 AAGCTGTGTACCTGAGTGGAGAAGAGCCCATCATCGCGCGGAGACCTTCTACCTGAC 1989  
DB 581 LysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrTyrTyrValAsp 600  
QY 1990 GCGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGGCTAGCTGACCGCGGCGCG 2049  
DB 601 GlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArg 620  
QY 2050 CAGAAGATCGTGAGCTGACCGAGACCAACCAAGAGAGACCGAGCTGCAGGCCATCCAG 2109  
DB 621 GlnLysIleValThrLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGln 640  
QY 2110 CTGGCCCTGCGAGGACGCGGAGCGAGGTGAACATCGTGCAGCGACGACGATCCGCTG 2169  
DB 641 LeuAlaLeuGlnAspSerGlySerGluValAlaAsnIleValThrAspSerGlnTyrAlaLeu 660  
QY 2170 GGCATCATCGAGCCCGACCGACAGAGCGAGCGAGCTGGTGAACACGATCATCCAG 2229  
DB 661 GlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleGlu 680  
QY 2230 CAGCTGTATCAAGAAGAGAAGTGTACCTGAGCTGGTGGCGCGCCACAAAGGCGCATCGGC 2289  
DB 681 GlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGly 700  
QY 2290 GGCACGAGCAGATCGACACAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTGTCTCTGAC 2349  
DB 701 GlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAsp 720  
QY 2350 GGCATCGAT 2358  
DB 721 GlyIleAsp 723  
RESULT 7  
ID AAE37601 standard; protein; 998 AA.  
XX AAE37601;  
XX 23-OCT-2003 (revised)  
DT 27-AUG-2003 (first entry)  
XX HIV-1 subtype C isolate Dui51 reverse transcriptase (RT) protein.  
KW Regulatory gene; accessory gene; HIV; human immunodeficiency virus;  
KW vaccine; infection; gene therapy; reverse transcriptase; RT; enzyme.  
XX Human immunodeficiency virus 1.  
XX WO2003037919-A2.  
XX 08-MAY-2003.  
XX 31-OCT-2002; 2002WO-IB004550.  
XX 31-OCT-2001; 2001ZA-00008978.  
XX (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.  
XX (UYCA-) UNIV CAPE TOWN.  
XX Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;  
XX

DR WPI: 2003-430497/40.  
DR N-PSDB; AAD29258.

XX New molecules comprising HIV-1 subtype isolate regulatory/accessory  
PT genes, useful for manufacturing a vaccine for treating or preventing HIV  
PT infection.

PS Disclosure; Page 94-97; 97pp; English.

XX The invention relates to molecules comprising HIV-1 subtype isolate  
CC regulatory/accessory genes (tat, nef and rev genes) and modifications and  
CC derivatives thereof. The invention also provides proteins encoded by such  
CC genes. Sequences of the invention are useful for manufacturing vaccines  
CC for treating or preventing human immunodeficiency virus (HIV) infections.  
CC They are also useful in gene therapy. The present sequence is HIV-1  
CC subtype C isolate Dnl51 reverse transcriptase (RT) protein. (Updated on  
CC 23-OCT-2003 to standardise OS field)

XX SQ Sequence 998 AA;

Alignment Scores:

Pred. No.: 1.4e-195 Length: 998  
Score: 3527.00 Matches: 662  
Percent Similarity: 96.51% Conservative: 29  
Best Local Similarity: 92.46% Mismatches: 21  
Query Match: 77.43% Indels: 4  
DB: 6 Gaps: 2

US-09-610-313B-32 (1-2457) x AAE37601 (1-998)

QY 223 TTCCGCGAGGACCTGGCTTCCCGAGGCGAGGCGCGAGTTCCTCCAGGAGCAGAAC 282  
DB 1 PheArgGluAenLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20  
QY 283 CGCCGCAACACCCACAGCGCGGAGCTGCAGGTGCGCGCGCAACACCCCGCGAGCGAG 342  
DB 21 ArgAlaenSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu 40  
QY 343 GCGCGCGCGAGCGCCAGGCGACCTGAACCTCCCGAGATCACCTGTGGCAGCGCCCC 402  
DB 41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60  
QY 403 CTGTGTGACATCAGGTGGGGCGGCAGATCAAGAGGCGCTGTGGACACCGGCGCGAC 462  
DB 61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAseThrGlyAlaAse 80  
QY 463 GACACCTGTGTGGAGGAGATGAGCTGCCGCGAAGTGGAGCCCAAGATGATCGCGGC 522  
DB 81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100  
QY 523 ATCGCGCGCTTCATCAAGGTGCGCGCAGTACACAGATCTGTGATCGAGATCTCGCGCAAG 582  
DB 101 IleGlyGlyPheIleLysValArgGlnThrAspGlnIleLeuIleGluIleCysGlyLys 120  
QY 583 AAGGCCATCGCACCGTGTGATCGGCGCACCCCGGTGAACATCATCGGCGCGCAACATG 642  
DB 121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140  
QY 643 CTGACCCAGCTGGGCTGCACCTCACTCCCATCAGCCCATCGAGCGGTGCCCGTG 702  
DB 141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160  
QY 703 AAGCTGAAGCCCGCATCGAGCGCGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAG 762  
DB 161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180  
QY 763 ATCAAGCCCTGACCGCATCTGCGAGGAGATGAGAGGAGGCGCAAGATCACCAAGATC 822  
DB 181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200  
QY 823 GGCCCCGAGACCCCTACACACCCCGTGTGCCATCAGAGAGGAGGAGCAGCACCAAG 882  
DB 201 GlyProGluAsnProThrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

QY 883 TGGCGCAAGCTGTGTGAGCTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTG 942  
DB 221 TrpArgLysLeuValAspPheArgGluLeuAenLysArgThrGlnAspPheTrpGluVal 240  
QY 943 CAGCTGGGCTCCCGCCCGCGCGCTGAAGAAGAGAGAGCTGACCGTGTGGAC 1002  
DB 241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAse 260  
QY 1003 GTGGCGGACGCTACTTTCAGGCTGCCCTGACGAGGACTTCCGCAAGTACACCCGCTTC 1062  
DB 261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280  
QY 1063 ACCATCCCGACATCAACACAGACACCCCGCATCCGCTACCAAGTACAACGCTGTGCC 1122  
DB 281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300  
QY 1123 CAGGCTGGAGGCGAGCGCGCGCTTCAGAGCAGCATGACCAAGATCTGTGGAGGCC 1182  
DB 301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320  
QY 1183 TTCCGCGCGCGCAACCCCGAGATCTGTATCTACACAG-----GCCCGCTGTACCTGGCG 1236  
DB 321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340  
QY 1237 AGCGACCTGGAGATCGCGCGACACCGCGCAAGATCGAGGAGCTGCGCAAGCACCCTGCTG 1296  
DB 341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuAseGluHisLeuLeu 360  
QY 1297 CGCTGGGCTTCACACCCCGCGCAAGAGCAGACAGAGGAGCGCGCTTCTCTCTG----- 1350  
DB 361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380  
QY 1351 CCCATCGAGCTGACCCCGCGCAAGTGGACCTGCGAGCGCCATCGAGCTGCCCGAGAGGAG 1410  
DB 381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400  
QY 1411 AGCTGGACCGTGAACGACATCCAGAGCTGTGTGGCGAAGCTGAACCTGGCGCGAGCCAGATC 1470  
DB 401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAenTrpAlaSerGlnIle 420  
QY 1471 TACCCCGCGCTCAAGGTGCGCGAGCTGTGAGCTGTGCGCGCGCGCGCAAGCGCTGACC 1530  
DB 421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440  
QY 1531 GACATCGTGGCCCTGACCGAGGAGCGAGCTGAGCTGGCGAGAGCCGCGAGATCTGTG 1590  
DB 441 AspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeu 460  
QY 1591 CGCGAGCGCGTGCACGCGCTGTACTACGACCCCGAGCAAGGACCTGTGTGGCGAGATCCAG 1650  
DB 461 LysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGln 480  
QY 1651 AAGCAGGCGCACGACGAGTGGAGCTTACAGATCTACAGAGCGCTTCAAGAACCTGAAG 1710  
DB 481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500  
QY 1711 ACCGGCAAGTACCGCAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1770  
DB 501 ThrGlyLysTyrAlaLysArgArgThrHisThrAsnAspValLysGlnLeuThrGlu 520  
QY 1771 GCCTGTGAGAAGATCGCATCGAGAGCATCGTGTATCTGGGCGAAGACCCCAAGTCTCGCG 1830  
DB 521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540  
QY 1831 CTGCCCATCCAGAGGAGACCTGGGAGACTGTGTGACCGACTACTGTGCGAGCGCACCTGG 1890  
DB 541 LeuProIleGlnLysGluThrTrpGluIleTrpThrAspTyrTrpGlnAlaThrTrp 560  
QY 1891 ATCCCGGAGTGGAGTTCGTGAACACCCCGCTGTGTGAGCTGTGTGTGTGTGTGTGTGTGT 1950  
DB 561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTyrTrpGlnLeuGlu 580



QY 723 CGGCCCCAAGGTGAAGCAAGTGGCCCTTCAACGAGGAGAGATCAAGGCCCTTGACCGCCCAT 782  
Db 654 pGlyProlysVallysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIle 674  
QY 783 CTGGGAGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACAA 842  
Db 674 eCyAspGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAenProTyrAS 694  
QY 843 CACCCCGTGTTCGCATCAAG 902  
Db 694 nThrProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPh 714  
QY 903 CCGGAGCTGAACAGCGCACCCAGGACTTCTGGAGAGTGAGCTGGGAGATCCCCCAACC 962  
Db 714 eArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734  
QY 963 CGCGCGCTGAAG 1022  
Db 734 oAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754  
QY 1023 CGTGCCTCGACAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAA 1082  
Db 754 rIleProLeuThrGluAspPheArgLysTyrThrAlaPheThrIleProSerArgAsnAs 774  
QY 1083 CGAGACCCCGGCATCCGCTACCAAGTACACAGCTGTCTGCCAGGCTGGAGGAGCGGCC 1142  
Db 774 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerLe 794  
QY 1143 CAGCATCTTCAGAGCAGCATGACAGATCCTGGAGCCCTTCCGCGCCCGCCACCCCGA 1202  
Db 794 uAlaIlePheGlnSerSerMetThrLysThrLeuGluProPheArgLysGlnAsnProGl 814  
QY 1203 GATCGTATCTACAG-----GCCCGCTCTAGCTGGCGAGCGACCTGGAGATCGGCA 1256  
Db 814 yIleValIleThrGlnTyrMetAspLeuThrValGlySerAspLeuGluIleGlyL 834  
QY 1257 GCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGCTTTCACCAACCC 1316  
Db 834 nHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrPr 854  
QY 1317 CGACAAGAGACACAG 1370  
Db 854 oAspLysLysHis---LysGluProProPheLeuThrMetGlyTyrGluLeuHisProAs 873  
QY 1371 CAAGTGGAGCTGCAGAGCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1430  
Db 873 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerTrpThrValAsnAspIl 893  
QY 1431 CCAGAAGCTGTGGGCAAGCTGAAGTGGCCAGCCAGATCTACCCCGGCATCAAGGTGCG 1490  
Db 893 eGlnLysLeuValGlyLysLeuAenTrpAlaSerGlnIleTyrProGlyIleLysValAr 913  
QY 1491 CCAGCTGTGAAGCTGTGCGCGGCGGCAAGCCCTGACCGAGATCGTGGCCCTTGACCGCA 1550  
Db 913 gGlnLysCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGl 933  
QY 1551 GGAGCGGAGCTGAGCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1610  
Db 933 uGluAlaGluLeuGluLeuAlaGluAenArgGluIleLeuLysGluProValHisGlyVa 953  
QY 1611 GTACTAGACCCAGCAG 1670  
Db 953 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973  
QY 1671 GACTTACAGATCTACAG 1730  
Db 973 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993  
QY 1731 CGGACCGCCGACACAG 1790  
Db 993 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

QY 1791 GGAGAGCATCTGTATCTGGGGCAAGAGACCCCAAGTTCCGCCTGCCATCCAGAGAGAGAC 1850  
Db 1013 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033  
QY 1851 CTGGGAGACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1910  
Db 1033 rTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053  
QY 1911 GAACACACCCCGCTGTGTGAAGCTGTGTGATCCAGCTGGAGAGAGGAGCCATCATCGGCG 1970  
Db 1053 lAsnThrProProLeuValLysLeuTyrTyrGlnLeuGluLysAspProIleValGlyVa 1073  
QY 1971 CGAGACCTTCTACCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 2030  
Db 1073 lGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTy 1093  
QY 2031 CGTCACACGACCGGCGCGCGCGAGAGATCGTGAGCTGACCGAGACCCACCAACCAAGACAC 2090  
Db 1093 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113  
QY 2091 CGAGCTCGAGGCCATTCAGCTGGCGCTGCAGGACAGCGGCGCGCGCGCGCGCGCGCTG 2150  
Db 1113 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133  
QY 2151 CGACAGCAGTACCGCTGGGCGATCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGAGCT 2210  
Db 1133 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153  
QY 2211 GGTGAACAGATCATCAGCAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2270  
Db 1153 uValAsnGlnIleIleGluGlnLeuMetLysLysGluArgValTyrLeuSerTrpValPr 1173  
QY 2271 CGCCCAACAGGCGATCGCGCGCAACAGAGCAGATCGACAAGCTGTGTGAGCAAGGCGATCCG 2330  
Db 1173 oAlaHisGlyGlyIleGlyLysGlnGluValAspLysLeuValSerSerGlyIleAr 1193  
QY 2331 CAAGCTGCTGTCTCGACCGCGCATCGAT 2358  
Db 1193 gLysValLeuPheLeuAspGlyIleAsp 1202  
RESULT 9  
AAM48949  
ID AAM48949 standard; protein; 854 AA.  
XX  
AC AAM48949;  
XX  
DT 29-AUG-2003 (revised)  
DT 19-APR-2002 (first entry)  
XX  
DE HIV-1 subtype C isolate Dul51 pol protein.  
XX  
KW HIV-1 subtype C; vaccine; HIV infection; AIDS; pol; antiviral.  
XX  
OS Human immunodeficiency virus; type I.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 58  
FT Misc-difference 72 /note= "encoded by TGA"  
FT Misc-difference 72 /note= "encoded by TAA"  
FT Misc-difference 114 /note= "encoded by TAA"  
FT Misc-difference 117 /note= "encoded by TAA"  
XX  
PD WO200204494-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 09-JUL-2001; 2001WO-IB001208.  
XX  
PR 07-JUL-2000; 2000US-0216995P.  
PR 10-JUL-2000; 2000ZA-00003437.



PR 15-SEP-2000; 2000ZA-00004924.  
 XX (MEDI-) MEDICAL RES COUNCIL.  
 PA (UYCA-) UNIV CAPE TOWN.  
 XX (UYNC-) UNIV NORTH CAROLINA.  
 PI Williamson C, Swanstrom RI, Morris L, Karim SA, Johnston RE;  
 XX WPI; 2002-171700/22.  
 DR N-PSDB; AAL41592.  
 XX  
 PT Selecting HIV-1 subtype C isolates, which are useful in developing  
 PT vaccines against HIV infection, comprises isolating viruses with high  
 PT sequence identity to a consensus sequence whose phenotype is associated  
 PT with the HIV subtype.  
 XX  
 PS Claim 19; Page 67; 69pp; English.  
 XX  
 CC The present invention relates to a process for the selection of human  
 CC immunodeficiency virus (HIV) subtype isolates for use in the development  
 CC of a prophylactic and/or therapeutic pharmaceutical composition. The  
 CC process involves selecting isolated virus or viruses with a high sequence  
 CC identity to a consensus sequence and a phenotype which is associated with  
 CC transmission for the particular HIV subtype. The composition can be used  
 CC in the production of vaccines against HIV. The present sequence is the  
 CC HIV-1 subtype C isolate Dui51 pol protein. (Updated on 29-AUG-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 854 AA;

Alignment Scores:  
 Pred. No.: 9,24e-193 Length: 854  
 Score: 3478.00 Matches: 662  
 Percent Similarity: 93.77% Conservative: 30  
 Best Local Similarity: 89.70% Mismatches: 31  
 Query Match: 76.36% Indels: 15  
 DB: 5 Gaps: 4

US-09-610-313B-32 (1-2457) x AAM48949 (1-854)

QY 171 CGGCAAGGAGGCGCCACGATGAAGACTGACACGAGCGCCAGGCCAATCTTCTT----- 224  
 Db 118 ArgGlnGlyPheProSerHisAspValValLysArgArgProValProSerLeuHisAla 137  
 QY 225 -----CCGCGAGGACTGGCC---TTCCCCCGAGGCGAAGCCCG 260  
 Db 138 CysArgSerThrLeuGluAspProArgValPro-SerSerPheProGlnGlyProAlaAr 157  
 QY 261 CGAGTTCCCGACGAGCAGAACCGCGCCACACGCCACCGCCGAGCTGCAGGTGG 320  
 Db 157 GginPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeuGlnValAr 177  
 QY 321 CGGCGACAAACCCCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380  
 Db 177 gArgAspAsnProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGl 197  
 QY 381 GATCACCCTGTGGACGCGCCCTGTGTAGCATCAAGTGTGGCGCGCGCGCGCGCGCGCG 440  
 Db 197 nileThrLeuTrpGlnArgProLeuValSerileLysileGlyGlyGlnThrArgGluAl 217  
 QY 441 CTTGTGGACACCG 500  
 Db 217 aleuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTr 237  
 QY 501 GAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGTGGCGCGCGCGCGCGCG 560  
 Db 237 pLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnThrAspGlnI 257  
 QY 561 CTGTATCGAGATCTCGCGCAAGGCCATCGGCGCCGCGCGCGCGCGCGCGCGCGCGCG 620  
 Db 257 eleuileGluileCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVa 277  
 QY 621 GAACATCATCGCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCCATCAG 680

Db 277 lAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSe 297  
 QY 681 CCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGCGATGAGCGGCCCAAGGTGAAGCA 740  
 Db 297 rProileGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGl 317  
 QY 741 GTGGCCCTGACCGAGGAGAGATCAAGCCCTGACCGCCATCTGCGAGGAGATGAGAA 800  
 Db 317 nTrpProLeuThrGluValLysIleLysAlaLeuThrAlaIleCysGluGluMetGluL 337  
 QY 801 GGAGGCAAGATCAACAGATCGGCCCGCGAAGAACCCCTACACACCCCGCTGTCGCCAT 860  
 Db 337 sGluGlyLysIleThrLysIleGlyProGluAsnProTyRAsnThrProIlePheAlaI 357  
 QY 861 CAAGAAGAACGACAGCACCAAGTGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAG 920  
 Db 357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysAr 377  
 QY 921 CACCCAGACTTCTGGGAGGTGCAGCTGGGCGATCCCCACCCCGCGCGCTGAAGAAGAA 980  
 Db 377 gThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLys 397  
 QY 981 GAAGAGCGTGCACCGTCTGCGACGCGCGCGCGCTTCTTACAGCGTGCCTCGACGAGCA 1040  
 Db 397 sLysSerValThrValLeuAspValGlyAspAlaTyRPheserValProLeuAspGlu 417  
 QY 1041 CTTCCGCAAGTACACCGCTTTCACATCCCCAGCATCAACACGAGACCCCGCGGATCGG 1100  
 Db 417 yPheArgLysTyRThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleAr 437  
 QY 1101 CTACCGTGTACACGCTGCTGCCCGCGGCTGGAAGGCGAGCCCGCGCTTCCAGAGAG 1160  
 Db 437 gTyRGlnTyRAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSe 457  
 QY 1161 CATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCGTGATCTACAG-- 1218  
 Db 457 rMetThrLysIleLeuGluProPheArgAlaLysAsnProGluIleValIleTyRGlnTy 477  
 QY 1219 ----GCCCCCTGTAGTGGCGAGCGACTGGAGATCGGCGAGCACCGCGCCCAAGATCGA 1274  
 Db 477 rMetAlaAlaLeuTyRValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGl 497  
 QY 1275 GGAGTGTGGCAAGCACCTGCTGCGCTGGGGCTTACACACCCCGCGACAGAGACCCAGAA 1334  
 Db 497 uGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnL 517  
 QY 1335 GGAGCCCGCTTCTCTG-----CCCATCGAGCTGCACCCCGCACAGTAGTGAGCGTCAGCC 1388  
 Db 517 sGluProProPheLeuTrpMetGlyTyRLeuLeuHisProAspLysTrpThrValGlnPr 537  
 QY 1389 CATCGAGTGTCCCGAAGAGGAGAGTGCACCGTGAAACGACATCCAGAGCTGGTGGGCA 1448  
 Db 537 oileGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyL 557  
 QY 1449 GCTGAGCTGGCGCGAGCATCTACCCCGCGCATCAAGTGTGGCGCGCGCTGCAAGTGT 1508  
 Db 557 sLeuAsnTrpThrSerGlnIleTyRProGlyIleLysValArgGlnLeuCysLysLeu 577  
 QY 1509 GCGCGCGCCCAAGGCGCTGACCGCATCTGTCGCCCTGACCGAGGAGGCGCGAGCTGAGCT 1568  
 Db 577 uArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGlu 597  
 QY 1569 GGCCGAGAACCGCGAGATCTTCGCGAGCCCGCTGTGACCGCGTGTACTACGACCCAGCA 1628  
 Db 597 uAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyRtyrAspProSerL 617  
 QY 1629 GGACCTGTGTGGCGAGATCCAGAGCGGCGCCACGACGAGTGGAGCTACCATCTACCA 1688  
 Db 617 sAspLeuIleAlaGluIleGlnLysGlnGlyAspAspGlnTrpThrTyRGlnIleTyRGl 637  
 QY 1689 GGAGCGCTTCAAGAACCTTGAAGACCGCGCAAGTACGCCAAGATGCGCACCGCGCCACCA 1748

Db 637 nGluProPheLysAsnLeuLysThrGlyLysTyAlaLysArgArgThrHisThrAs 657  
 Qy 1749 CGACGTGAAGCAGCTGACGAGCGCGTCGACAGATCGCCATGAGAGCATCGTGATCTG 1808  
 Db 657 nAspValLysGlnLeuThrGluAlaValGlnLysIleSerLeuGluSerIleValThrTr 677  
 Qy 1809 GGGCAAGACCCCAAGTTCGCTCCGCTCCATCCAGAGAGACCTGGAGACCTGGTGGAC 1868  
 Db 677 pGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpTh 697  
 Qy 1869 CGACTACTGGCAGCGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCGCCCTGGGT 1928  
 Db 697 rAspTyrrProGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuVal 717  
 Qy 1929 GAAGCTGTGTACAGCTGGAGAGAGAGCCCATCATCGCGCCCGAGACCTTCTACGTGGA 1988  
 Db 717 LysLeuTrpTyrrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPheTyrrValAs 737  
 Qy 1989 CGGCGCGCCGACCGCAGACCAAGATCGGCAAGCGCGGTACCTGACCGACCGGGCCG 2048  
 Db 737 pGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrrValThrAspArgGlyAr 757  
 Qy 2049 GCAGAAGATCGTGAGCTGACCGAGACCCACCAACAGAGACCGAGCTGCAGGCGCATCCA 2108  
 Db 757 gGlnLysIleValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGl 777  
 Qy 2109 GCTGCGCCTCGCAGACAGCGCAGAGGTGAACATCGTGACCCAGACGACGAGTACGCGCT 2168  
 Db 777 nLeuAlaLeuGlnAspSerGluSerGluValAsnIleValThrAspSerGlnTyrrAlaLe 797  
 Qy 2169 GGGCATCATCAGCGCCGAGCGCCGACAGAGCGAGAGCGAGCTGTGTGAACAGATCATCGA 2228  
 Db 797 uGlyIleGlnAlaGlnProAspArgSerGluSerGluLeuValAsnGlnIleIleGl 817  
 Qy 2229 GCAGCTCATCAAGAGAGAGGTGTACCTGAGCTGGTGGTCCCGCCACAGAGGCGATCGG 2288  
 Db 817 uGlnLeuIleLysLysGluArgAlaTyrrLeuSerTrpValProAlaHisLysGlyIleGl 837  
 Qy 2289 CGGCAACGAGCAGATCAGACAGCTGTGTGAGCAAGGCGCATCGCAAGGTGTGTG 2340  
 Db 837 yGlyAspGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 854

RESULT 10  
 AAB69282  
 ID AAB69282 standard; protein; 1000 AA.  
 XX  
 AC AAB69282;  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 XX HIV-1 non-subtype B clone 92RW009-6 pol protein.  
 XX  
 DE HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 KW  
 XX Human immunodeficiency virus 1.  
 OS  
 XX  
 XX W0200026416-A1.  
 XX  
 XX 11-MAY-2000.  
 XX  
 XX 25-OCT-1999; 99WO-US024837.  
 XX  
 XX 02-NOV-1998; 98US-00184418.  
 PR  
 XX (UABR-) UAB RES FOUND.  
 PA  
 XX Hahn BH, Shaw GM, Gao P;  
 PI  
 XX WFI; 2000-365651/31.  
 DR  
 XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.

PS Claim 41; Fig 15; 131pp; English.

XX The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 1000 AA;

Alignment Scores:  
 Pred. No.: 1.52e-192 Length: 1000  
 Score: 3474.50 Matches: 656  
 Percent Similarity: 95.40% Conservative: 29  
 Best Local Similarity: 91.36% Mismatches: 28  
 Query Match: 76.28% Indels: 5  
 DB: 3 Gaps: 3

US-09-610-313B-32 (1-2457) x AAB69282 (1-1000)

Qy 220 TTCTTCGCGAGACCTGGCTTCCCCAGGCGAGCCCGCGAGTTCCCGAGCGAGCAG 279

Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20

Qy 280 AACCGCGCCCAACGCCACCCAGCGCGAGCTGCAGTGCAGCGGC---GACAAACCCCGC 336

Db 21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyGlyArgAspSerLeuSer 40

Qy 337 AGCGAGCGCGCGCGCGAGCGCGAGCGCACCTGACCTTCCCGAGATCACCCTGTGCAG 396

Db 41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnIleThrLeuTrpGln 60

Qy 397 CGCCCTGTGTGAGCATCAAGTGGCGCGCGAGCGCGAGCGCGCTGTGCAGCACCGGC 456

Db 61 ArgProLeuValThrValLysIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80

Qy 457 GCCGACGACACCGTGTGTGAGGAGATGAGCTGCGCGCGAGTGAAGCCCAAGATGATC 516

Db 81 AlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetile 100

Qy 517 GCGGCGATCGCGCGCTTCATCAAGTGCAGCGCGAGTGCAGCACGATCCTGATCGATCTGC 576

Db 101 GlyGlyIleGlyGlyPheIleLysValLysGlnTyrrAspGlnIleLeuIleGluIleCys 120

Qy 577 GGCAAGAAGCGCATCGGCGCGCTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 636

Db 121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleIleGlyArg 140

Qy 637 AACATGCTGACCGAGCTGGGCTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTG 696

Db 141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160

Qy 697 CCGGTGAGCTGAAGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 756

Db 161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 180

Qy 757 GAGAAGATCAAGCGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 816

Db 181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluGlyLysIleSer 200

Qy 817 AAGATCGCGCGCGAGAACCCCTACACACCCCGTGTTCGCCCATCAAGAGAGAGCAGC 876

Db 201 LysIleGlyProGluAsnProTyrrAsnThrProValPheAlaIleLysLysLysAspSer 220

Qy 877 ACCAAGTGGCGCAAGCTGTGTGAGCTTCCCGAGCTGAAACAGCGCACCCAGGACTTCTGG 936

Db 221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240

QY 937 GAGTGCAGCTGGGCGATCCCCACCCCGCCGCTGAAGAGAAAGAGCGTGCAGCTG 996  
Db 241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysValThrVal 260  
QY 997 CTGGACGTGGGCGAGCCCTACTTCAGCGTCCCTCGACGAGACTTCGCAAGTACAC 1056  
Db 261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280  
QY 1057 GCCTTACATCCCGCAGCATCAACACAGACCCCGGCATCCGTCACAGTACACCTG 1116  
Db 281 AlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 300  
QY 1117 CTGCCCCAGGCTGAAGGCGACCCAGCATCTCCAGAGCAGCATCAACAGATCCTG 1176  
Db 301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320  
QY 1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCTACCCAG- - - - -GCCCCCTGTAC 1230  
Db 321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspAspLeuTyr 340  
QY 1231 GTGGCAGCGACTCGGATCGGCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCAC 1290  
Db 341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHis 360  
QY 1291 CTGCTCGCTGGGCTTCACACCCCGCAGAGAGCAGCAGAGAGCCCGCTTCCTG 1350  
Db 361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 380  
QY 1351 - - - - -CCCATCGAGCTGCACCCCGACCAAGTGGACCGTGCAGCCCATCGAGCTGCCGAG 1404  
Db 381 TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400  
QY 1405 AAGGAGAGTGAACCGTGAACACATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCCAGC 1464  
Db 401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420  
QY 1465 CAGATCTACCCCGCATCAGGTGGCGCAGCTGTGCAAGCTGCTGGCGGCGCCAGGCC 1524  
Db 421 GlnIleTyrProGlyValLysValArgGlnLeuLysLysLeuLeuArgGlyThrLysAla 440  
QY 1525 CTGACCGCATCGTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGAGACCGCGAG 1584  
Db 441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGlu 460  
QY 1585 ATCTGGCGGAGCCGTCACCGCTGTACTAGACCCCGCAGCAAGGACTGTGTGGCCGAG 1644  
Db 461 IleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGlu 480  
QY 1645 ATCCAGACGAGCGCCAGCACGACGACCTACAGATCTACCGAGGCCCTTCAGAAC 1704  
Db 481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500  
QY 1705 CTGAAGACCGCAAGTACGCCAAGATGGCCACCGCCACCAACGAGCTGAAGCAGCTG 1764  
Db 501 LeuLysThrGlyLysTyrAlaLysArgThrAlaHisThrAsnAspValLysGlnLeu 520  
QY 1765 ACCGAGGCGCTGAGAAGATCCCATGGAGAGCATCGTGATCTGGGCAAGACCCCAAG 1824  
Db 521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrpGlyLysThrProLys 540  
QY 1825 TTCGCGCTGCCATCCAGAGAGACCTGGGAGACTGGTGACCGACTACTGCGAGGCC 1884  
Db 541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAla 560  
QY 1885 ACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTCGTGAAGCTGTGGTACCAG 1944  
Db 561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580  
QY 1945 CTGGAAGAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAGCGGCGCCCAACGCC 2004  
Db 581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

QY 2005 GAGCAAGATCGGCAAGCGCGCTACGTGACCGACCGCGGCGCGGCGAGAGATCGTGAGC 2064  
Db 601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620  
QY 2065 CTGACCGAGACCCACCAAGACCGAGGTGAGGCGCATCCAGCTGGCCCTGCAGGAC 2124  
Db 621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640  
QY 2125 AGCGCAGCAGGTGAACATCGTACCGACGACGACGAGTACGCGCTGGGCGATCCAGGCC 2184  
Db 641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 660  
QY 2185 CAGCCGCAAGAGAGGAGCGAGCGAGCTGGTGAACACAGATCATCGACGAGTGTATCAAGAAG 2244  
Db 661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleIleGluGlnLeuLysLys 680  
QY 2245 GAGAAGGTGTACCTGAGCTGGTGGTCCCGCCACCAAGGCGCATCGCGGCAACGAGCAGATC 2304  
Db 681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 700  
QY 2305 GACAAGCTGTGAGCAGGCGCATCGCAAGGTGCTGTCTGCGAGCGGCATCGAT 2358  
Db 701 AspLysLeuValSerSerGlyIleArgArgValLeuPheLeuAspGlyIleAsp 718  
RESULT 11  
AAW72993  
ID AAW72993 standard; protein; 1002 AA.  
XX  
AC AAW72993;  
XX  
DT 25-MAR-2003 (revised)  
DT 15-FEB-1999 (first entry)  
XX  
XX HIV isolate LAV.MAL pol protein.  
XX LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;  
KW pol.  
XX Human T-lymphotropic virus.  
XX  
XX US5824482-A.  
XX 20-OCT-1998.  
XX 06-JUN-1995; 95US-00471474.  
XX 23-JUN-1986; 86FR-00040138.  
PR 13-APR-1987; 87US-00038330.  
PR 19-FEB-1991; 91US-00656797.  
PR 10-DEC-1992; 92US-00988530.  
PR 18-NOV-1993; 93US-00154397.  
XX (INSP ) INST PASTEUR.  
XX Wain-Hobson S, Sonigo P, Alizon M, Montagnier L;  
XX WPI; 1998-582548/49.  
DR N-P5DB; AAV63467.  
XX Human immunodeficiency virus isolate LAV (MAL) - and method for detecting  
PT anti-HIV antibodies.  
XX Disclosure; Fig 7B-E; 47pp; English.  
XX This is the amino acid sequence of the pol protein of lymphadenopathy  
CC associated virus LAV.MAL (NCIM 1-641), a new virus isolate from Zaïre  
CC that is responsible for diseases clinically related to AIDS. The sequence  
CC was deduced from an open reading frame (ORF) of the LAV.MAL genome (see  
CC AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of  
CC the envelope glycoprotein can be used as antigens in a claimed method for  
CC the in vitro detection of an antibody directed against LAV. The method is  
CC useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in  
CC patients, asymptomatic carriers and in blood-related products. LAV viral

CC antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct  
CC PR field.)

XX Sequence 1002 AA;

Alignment Scores:  
Pred. No.: 1,68e-191 Length: 1002  
Score: 3456.50 Matches: 649  
Percent Similarity: 95.56% Conservative: 39  
Best Local Similarity: 90.14% Mismatches: 25  
Query Match: 75.88% Indels: 7  
DB: 2 Gaps: 4

US-09-610-313B-32 (1-2457) x AA#W2993 (1-1002)

Qy 220 TTCTTCCGCGAGGACCTGGCCCTTCCCGAGGCGAGGCGGCGGAGTTCCCGAGGCGAG 279  
Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerSerGluGln 20  
Qy 280 AACCGCGCAACACCCCGACCGCGGAGTGGAGTG---CGCGGCGACAAACCCCGCG 336  
Db 21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40  
Qy 337 AGCGAGCGCGCGCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 390  
Db 41 SerGluThrGlyAlaGluArgGlnGlyLeValSerPheSerPheProGlnThrLeu 60  
Qy 391 TGGCAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 450  
Db 61 TrpGlnArgProValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80  
Qy 451 ACCGCGCGCGACACCGCTGCTGGAGGAGATGAGCTGCGCGGCGGCGGCGGCGGCGGCGG 510  
Db 81 ThrGlyAlaAspAspThrValLeuGluGluLeuAsnLeuProGlyLysTrpLysProLys 100  
Qy 511 ATGATCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 570  
Db 101 MetIleGlyLysGlyGlyPheIleLysValArgGlnTrpAspGlnThrLeuLeuGlu 120  
Qy 571 ATCTCGCGCGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 630  
Db 121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleLeu 140  
Qy 631 GGCGCGCAACATGTCACCGCTGGCTGGCTGACCTGAACTTCCCGGCGGCGGCGGCGGCGG 690  
Db 141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
Qy 691 ACCGTGCGCGTGAAGCTGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 750  
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180  
Qy 751 ACCGAGGAGAGATCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 810  
Db 181 ThrGluGlyLysLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200  
Qy 811 ATCAACCAAGATCGCGCGCGGAGAACCCCTACAACACCGCGGCTGTTCGCGCATCAAGAAG 870  
Db 201 IleLeuLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLysLys 220  
Qy 871 GACAGCACCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 930  
Db 221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240  
Qy 931 TTCTGGGAGGTGGAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 990  
Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
Qy 991 ACCGTGCTGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1050  
Db 261 ThrValLeuAspValGlyAspAlaTyThrPheSerValProLeuAspGluAspPheArgLys 280  
Qy 1051 TACACGCGCTTACCATCCCGAGCATCAACAGAGACCGCGGCGGCGGCGGCGGCGGCGGCGG 1110

Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300  
Qy 1111 AACGTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1170  
Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
Qy 1171 ATCTTGAGGCGCTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1224  
Db 321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
Qy 1225 CTGTACGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1284  
Db 341 LeuTyrValGlySerAspLeuLysGlyGlnHisArgThrLysIleGluGluLeuArg 360  
Qy 1285 AAGCACCTGCTGCGCTGGGCGGCTTCAACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1344  
Db 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
Qy 1345 TTCTCTG-----CCCATCGAGCTGCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1398  
Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
Qy 1399 CCGGAGAGGAGGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGCGGCGGCGGCGGCGG 1458  
Db 401 ProAspLysGluSerTrpThrValAsnAspLysGlnLysLeuValGlyLysLeuAsnTrp 420  
Qy 1459 GCCAGCGCAGATCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1518  
Db 421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440  
Qy 1519 AAGGCGCTGACCGACATCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1578  
Db 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460  
Qy 1579 CGCGAGATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1638  
Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuLeu 480  
Qy 1639 GCCGAGATCCAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1698  
Db 481 AlaGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGlnTyr 500  
Qy 1699 AAGAACCTGAAGACCGGCAAGTACCGCAAGATCGCGGCGGCGGCGGCGGCGGCGGCGG 1758  
Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
Qy 1759 CAGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1818  
Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
Qy 1819 CCCAAGTTCCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1878  
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpThrThrGluTyrTrp 560  
Qy 1879 CAGGCGCACCTGGATCCCGGAGTGGGAGTTCGTGAACACCGCGGCGGCGGCGGCGGCGG 1938  
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
Qy 1939 TACAGCTGGAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1998  
Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
Qy 1999 AACCGCGGAGACCAAGTCCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2058  
Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
Qy 2059 GTGAGCTGACCGGAGACCAACCGAGAGACCGGAGCTGCGGCGGCGGCGGCGGCGGCGG 2118  
Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
Qy 2119 CAGGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2178  
Db 641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660







QY 568 GAGATCTGCGCAAGAAGCCATCGGCACCGTGTGATCGCGCCACCCCGCTGAACATC 627  
 Db 121 GluileCysGlyHisIysAlaileGlyThrValLeuValGlyProThrProValAsnile 140  
 QY 628 ATCGCGCGCAACATGCTGACCGAGCTGGGTGCGACCTGAACTTCCCATCGACCCCATC 687  
 Db 141 IleGlyArgAsnLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGGCCCAAGTGAAGCAGTGGCC 747  
 Db 161 GluThrValProValIysLeuLysProGlyMetAspGlyProLysValIysLeuThrPro 180  
 QY 748 CTGACCGAGGAGAGATCAAGCCCTGACCGCCATCTCGAGGAGATGAGAGAGGGC 807  
 Db 181 LeuThrGluLysIleIysAlaLeuValGluileCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCAACAAGATCGGCGCCGAGAACCCCTCAACACCCCGTTCGCCATCAAGAAG 867  
 Db 201 LysIleSerLysIleGlyProGluAsnProIleAsnThrProValPheAlaIleLysLys 220  
 QY 868 AAGGACAGCACAAGTGGCGGCAAGCTGGTGGACTTCCGGAGCTGAACAAGCGCACCCAG 927  
 Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
 QY 928 GACTTCTGGAGGTGCAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAAGAGAGC 987  
 Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGAGCGTGGCGACCGCTACTTACGCTGCCCTCGACGAGGACTTCCGC 1047  
 Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280  
 QY 1048 AAGTACACCGCTTACCATCCCGAGCATCAACAGAGACCCCGGATCCGCTACCGAG 1107  
 Db 281 LysIleThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
 QY 1108 TACAACGTGCTCCCCAGGCTGGAAGGCGACCGCCAGCATCTTCCAGAGCAGATGACC 1167  
 Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 QY 1168 AAGATCTGGAGCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACGAG-----GCC 1221  
 Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
 QY 1222 CCCCTGTAGTGGCAGCGACTGAGATCGCGCAGCACCGCCCAAGATCGAGGAGCTG 1281  
 Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 360  
 QY 1282 CGCAAGCACCTGCTGGCTGGGGCTTACACACCCCGCAAGAAGCAGAGAGAGGCC 1341  
 Db 361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCTCTG-----CCCATCGAGTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAG 1395  
 Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
 QY 1396 CTGCGCGAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAAC 1455  
 Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
 QY 1456 TGGCGCAGCAGATCTACCCCGGCATCAAGTGGCGCAGCTGTCAAGCTGTGCGCGCGC 1515  
 Db 421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
 QY 1516 GCCAAGGCCCTGACGACATCGTGCCTGACCGGAGGCGGAGCTGAGTGGCCGAG 1575  
 Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
 QY 1576 AACCCGAGATCCTGGCGGCGCCGTGCACCGCGCTGTACTACGACCCGAGGAGACCTG 1635  
 Db 461 AsnArgGluIleLysGluProValHisGlyValTyrAspProSerLysAspLeu 480  
 QY 1636 GTGGCGCAGATCCAGAAGCAGGGCCACGACCGAGTGGACCTACGAGTCTACCGAGGCC 1695

Db 481 IleAlaGluileGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500  
 QY 1696 TTCAAGACCTGAGACCGGCAAGTACCGCAAGATGCCACCGCCACCAACGACGCTG 1755  
 Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
 QY 1756 AAGCAGCTGACCGAGGCGCTGCAGAAAGATGCCATGGAGAGATCGTGTGATCTGGCGCAAG 1815  
 Db 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540  
 QY 1816 ACCCCAAAGTTCGCGCTGCCATCCAGAAAGAGACCTGGGAGACCTGTGTGACCACTAC 1875  
 Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTyrGluThrTyrTrpThrGluTyr 560  
 QY 1876 TGGCAGGCGACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGTGAGCTG 1935  
 Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 580  
 QY 1936 TGGTACCAGCTGGAGAGGAGCCCATCATCGCGCGGAGACCTTCTAGCTGGACGCGCC 1995  
 Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
 QY 1996 GCCAACCCGAGACCAAGATCGGCAAGCGCGCTACGTCACCGACCGCGCGCGGCAAG 2055  
 Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620  
 QY 2056 ATCGTAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115  
 Db 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640  
 QY 2116 CTGACGAGACCGCGCAGCGAGGTGAACATCGTACCGACAGCAGCAGTACGCTGGGCATC 2175  
 Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
 QY 2176 ATCCAGGCGCCAGCCGACAGAGCGAGCGAGCTGGTGGTGAACACAGATCATCGAGCAGCTG 2235  
 Db 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGlnLeu 680  
 QY 2236 ATCAAGAGGAGAGAGTGTACTGAGCTGGGTGGCGCCGCCACAGGGCATCGCGCGCAC 2295  
 Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700  
 QY 2296 GAGCAGATCGCAAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTGTCTCTGACGCGCATC 2355  
 Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
 QY 2356 GAT 2358  
 Db 721 Asp 721  
 RESULT 14  
 ID AAP81861 standard; protein; 1002 AA.  
 AC AAP81861;  
 XX 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 16-DEC-1990 (first entry)  
 XX Sequence encoded by LAV MA L POL gene.  
 XX HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.  
 XX Lymphadenopathy-associated virus.  
 XX W08707906-A.  
 XX 30-DEC-1987.  
 XX 22-JUN-1987; 87WO-EP000326.  
 XX



Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
Qy 1819 CCCAGTTCCGCTCCCATCAGAGGAGACCTGGGACACTGTGGACCGACTACTGG 1878  
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTrpTrp 560  
Qy 1879 CAGGCCACTGTGATCCCGAGTGGGAGTTCGTGAACACCCCTGTGTGAAGCTGG 1938  
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
Qy 1939 TACCAGCTGGAGAAGGCCCATCATCGCGCCGAGACCTTCTAGCTGGACGCGCCGCC 1998  
Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyr-ValAspGlyAlaIle 600  
Qy 1999 AACCCGAGACCAAGATCGGACCGCGCTACGTACCGACCGGGGGCGGAGAGATC 2058  
Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgLysVal 620  
Qy 2059 GTGAGCCTGACCGAGACCAACACAGAGACCGAGCTGCAGGCCATCCAGCTGCCCTG 2118  
Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
Qy 2119 CAGGACAGCGGACGAGGTGAACATCGTGACCGACGACGACGACGACGACGACG 2178  
Db 641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
Qy 2179 CAGGCCCGCCGACAGGACGAGCGAGCTGGTGAACACGATCATCGACGAGCTGATC 2238  
Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluIleLeuIle 680  
Qy 2239 AAGAAGGAGAGGTGTACTGAGCTGGGTGCCGCCCAAGAGGATCGCGGCAACGAG 2298  
Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyAsnGlu 700  
Qy 2299 CAGATCGAACGCTGGTGAAGGCGCATCGCAAGGTGCTGTCTCTGACGCGCATCGAT 2358  
Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 15  
ABR55489  
ID ABR55489 standard; protein; 1003 AA.  
AC ABR55489;  
XX  
DT 11-AUG-2003 (first entry)  
XX  
DE Amino acid sequence of a HIV pol protein.  
XX  
KW Polymorphic marker; host response; HIV; AIDS related virus; vaccine;  
KW hepatitis related virus; HCV; HBV; drug resistance; pol.  
XX  
OS Human immunodeficiency virus.  
XX  
PN WO2003035097-A1.  
XX  
PD 01-MAY-2003.  
XX  
PF 23-OCT-2002; 2002WO-AU001450.  
XX  
PR 23-OCT-2001; 2001AU-00008425.  
XX  
PA (EPIP-) EPIPOP PTY LTD.  
XX  
PI Mallal S;  
XX  
DR WPI; 2003-449231/42.  
XX  
PT Determining the influence of variation in host genes on the selection of  
PT microorganisms with protein substitutions, comprises typing individuals  
PT of a cohort infected with a microorganism for an intrinsic polymorphic  
PT marker.  
XX  
PS Claim 22; Page 88-89; 157pp; English.

XX The specification describes a method of determining the influence of  
CC variation in host genes on selection of microorganisms with protein  
CC substitutions. The method comprises typing all individuals of a  
CC population of patients infected with a microorganism for at least one  
CC selected intrinsic polymorphic marker involved in the host response to  
CC the presence of the microorganism. The method is useful for examining  
CC selective pressures confronting a wide range of organisms that exhibit  
CC pathogenic traits in a host, such as bacteria, fungi, mycobacterium,  
CC viruses and virus-like particles; for examining microorganisms that have  
CC adapted to evolve rapidly, including HIV and AIDS related viruses and the  
CC hepatitis related viruses such as HCV and HBV. The method is useful for  
CC designing a vaccine to prevent or delay the emergence of drug resistance  
CC in patients treated with a particular drug specific for a microorganism  
CC where the drug affects the replication of the microorganism at the  
CC nucleotide or amino acid level. The present sequence represents a HIV pol  
CC protein, which is expected to provide optimal cytotoxic T lymphocyte  
CC (CTL) induced therapeutic protection to the cohort examined in that study  
XX  
SQ Sequence 1003 AA;  
Alignment Scores:  
Pred. No.: 3,49e-191 Length: 1003  
Score: 3451.00 Matches: 650  
Percent Similarity: 95.15% Conservative: 36  
Best Local Similarity: 90.15% Mismatches: 27  
Query Match: 75.76% Indels: 8  
DB: 6 Gaps: 4  
US-09-610-313B-32 (1-2457) x ABR55489 (1-1003)  
Qy 220 TTCCTCCGAGGACCTGGCTTCCCCAGGCGAAGGCCGCGAGTCTCCCGAGCGAGCAG 279  
Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20  
Qy 280 AACCGCGCAACAGACCCACCAGCGCGAGCTGCAGGTGCGCGGCGAC-----AACCCC 333  
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGluAspAsnSer 40  
Qy 334 CGCAGCGAGCGCGCGCGCGAGCGCGCGAGCGCGACCTCTG-----AAGTCTCCCGAGTACC 387  
Db 41 ThrSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60  
Qy 388 CTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGCGGCGCGAGTCAAGAGGCGCCCTGTG 447  
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
Qy 448 GACACCGCGCGCGCGACACACCTGTCTGGAGGAGATGAGCTCCCGCGCAAGTGGAGGCC 507  
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100  
Qy 508 AAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGGCGCGAGTACGACAGATCTGTATC 567  
Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleIleIle 120  
Qy 568 GAGATCTGCGCGCAAGAGCGCCATCGGACCTGTGATCGCGCGCGCGCGCGCGCGCGCATC 627  
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
Qy 628 ATCGCGCGCAACATGTCTGACCCAGCTGGGCTGCGACCTCGAACTTCTCCCATCAGCCCCATC 687  
Db 141 IleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIle 160  
Qy 688 GAGACCGTGGCGTGAAGCTGAAGCCCGGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCC 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
Qy 748 CTGACCGAGGAGAGATCAAGCCCTGACCGCCATCTCGAGGAGATCGAGAGAGGCGC 807  
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
Qy 808 AAGATCAACAAAGATCGCGCGCGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAG 867

201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220 Db  
968 AAGCAGACACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTTGAACAAGCGCACCCAG 927 Qy  
221 LysAspSerThrLysTyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240 Db  
928 GACTCTCTGGAGGTGCAGCTGGGATCCCCACCCCGCGGCTGAGAGAAGAAGAGC 987 Qy  
241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260 Db  
988 GTGACCGTCTGGACGTGGGGACGCTTACTTTCAGCGTCCCTGGACGAGGACTCCGC 1047 Qy  
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280 Db  
1048 AAGTACACCGCTTACCACATCCCGAGCATCAACACGAGACCCCGCGCATCCGCTACCAG 1107 Qy  
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300 Db  
1108 TAGAACGTGTGCCCCAGGGCTGAAGGGAGCCCCAGCATCTTCAGAGCAGCATGACC 1167 Qy  
301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320 Db  
1168 AAGATCTCTGGAGCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCCAG-----GCC 1221 Qy  
321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340 Db  
1222 CCCCTGTACGTGGCGCAGCGACTCGAGATCGCCGAGCACCGCGCAAGATCGAGAGCTG 1281 Qy  
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360 Db  
1282 CGCAAGCACCTGTGCGCTGGGGCTTCAACACCCCGACAGAAGACCCAGAGAGGCC 1341 Qy  
361 ArgGlnHisLeuLeuLysTrpGlyPheThrProAspLysLysHisGlnLysGluPro 380 Db  
1342 CCCTTCTCG-----CCCATCGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAG 1395 Qy  
381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400 Db  
1396 CTGCCCGAGAGGAGAGCTGGACCGTGAACACATCCAGAACGTGGTGGGCAAGCTGAAC 1455 Qy  
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420 Db  
1456 TGGCCAGCCAGATCTACCCCGGATCAAGTGGCGCAGCTGTGCAAGCTGTGCGCGGC 1515 Qy  
421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440 Db  
1516 GCCAAGGCCCTGACCGACATGTCCTCCCTGACCGAGGCGCGAGCTGGAGCTGGCCGAG 1575 Qy  
441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460 Db  
1576 AACCGCAGATCTCTGCGCAGCGCTGCACGGCTGTACTACGACCCCGACGAGGACCTG 1635 Qy  
461 AsnArgGluLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480 Db  
1636 GTGCCCGAGATCCAGAGAGGGCCACGACAGTGGACCTTACCAGATCTTACCAGAGCCCC 1695 Qy  
481 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500 Db  
1696 TTCAGAACCTGAGACCGCGCAGTAGCCAGATGGCCACCGCCGCCACACCAAGCAGCTG 1755 Qy  
501 PheLysAsnLeuTyrThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520 Db  
1756 AAGCAGCTGACCGAGGCGGTGCAGAAGATCGCATGGAGAGCATCGTGATCTGGGGCAAG 1815 Qy  
521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540 Db  
1816 ACCCCCAGTTCGCCCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875 Qy  
541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTyrTrpThrGluTyr 560 Db  
1876 TGGCAGGCCACCTGGATCCCGGAGTGGAGTTCGTGAACACCCCGCCCTGGTGGAGCTG 1935 Qy  
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580 Db

1936 TGTATCCAGCTGGAGAAAGGAGCCCATCATCGGCCCGAGACCTTCTACGTGAGCGGCC 1995 Qy  
581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600 Db  
1996 GCCAACCGCAGAGCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGGCGCGCAGAG 2055 Qy  
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620 Db  
2056 ATCTGTAGCTGACCGAGACCCCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCC 2115 Qy  
621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640 Db  
2116 CTGACGACGACGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCTGGGCATC 2175 Qy  
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660 Db  
2176 ATCCAGGCCCGCAGCCGACAGAGCGAGAGCTGTGAACACGATCATCGACGAGCTG 2235 Qy  
661 IleGlnAlaGlnProAspLysSerGluLeuValSerGlnIleIleGluGlnLeu 680 Db  
2236 ATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCCACAGGGCATCGCGGCAAC 2295 Qy  
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700 Db  
2296 GAGCAGATCGACAGCTGTGTGACGAGCGGATCCGCAAGGTGTCTTCTGAGACGGCATC 2355 Qy  
701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720 Db  
2356 GAT 2358 Qy  
721 Asp 721 Db

Search completed: June 2, 2005, 04:43:53  
Job time : 233.424 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 02:53:39 ; Search time 40.7339 Seconds  
(without alignments)

11607.276 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 4555

Sequence: 1 gtcagccaccatg9ccga.....gggctagcaccggtgaattc 2457

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09610313/runat\_31052005\_155137\_15159/app\_query\_faeta\_1.7893  
-DB=PIR\_79 -QFWT=fastan -SURFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=BITS -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09610313 @CGN 1 1 189 @runat\_31052005\_155137\_15159 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR\_79:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3462.5	76.0	1002	2	S54378 pol polyprotein -
2	3454.5	75.8	1002	1	GNLJND HIV-1 retropepsin
3	3449	75.7	1003	1	GNVWIV HIV-1 retropepsin
4	3432	75.3	1015	1	GNVWH3 HIV-1 retropepsin
5	3410	74.9	1003	2	T09440 pol polyprotein -
6	3407	74.8	1003	1	B44001 HIV-1 retropepsin
7	3406	74.8	1003	1	GNVW2 HIV-1 retropepsin
8	3396	74.6	1012	1	GNVWVL HIV-1 retropepsin
9	3168	69.5	1027	1	GNLJST HIV-1 retropepsin
10	3064	67.3	912	2	S33980 pol polyprotein -
11	3051	67.0	902	2	T01668 pol polyprotein -
12	2728	59.9	559	2	B47175 reverse transcript
13	2708	59.5	559	2	A47175 reverse transcript
14	2307	50.6	1039	2	S46347 pol polyprotein -

15	2170	47.6	1055	2	S53092 pol polyprotein -
16	2166.5	47.6	1055	1	GNLJST HIV-1 retropepsin
17	2158	47.4	1036	1	GNLJG2 HIV-1 retropepsin
18	2157.5	47.4	1056	1	GNLJG3 HIV-1 retropepsin
19	2149.5	47.2	1035	1	GNLJGG HIV-1 retropepsin
20	2149	47.2	1034	1	GNLJCA HIV-1 retropepsin
21	2143.5	47.1	1019	2	T11560 pol polyprotein -
22	2130.5	46.8	1009	2	S28081 pol polyprotein -
23	2111.5	46.4	1032	2	S12153 pol polyprotein -
24	2034	44.7	1058	2	S08436 pol polyprotein -
25	2015	44.2	1061	1	GNLJG4 HIV-1 retropepsin
26	1980.5	43.5	1054	1	GNLJG5 HIV-1 retropepsin
27	1925.5	42.3	656	2	S30484 pol polyprotein -
28	1896.5	41.6	656	2	S30483 pol polyprotein -
29	1350	29.6	1124	2	GNLJFP HIV-1 retropepsin
30	1347	29.6	1124	1	GNLJFF HIV-1 retropepsin
31	1305	28.6	1124	2	B45557 HIV-1 retropepsin
32	1122.5	24.6	1109	1	B45345 HIV-1 retropepsin
33	1106.5	24.3	1087	2	JQ1162 pol protein - Maed
34	1102.5	24.2	1086	1	B46335 HIV-1 retropepsin
35	1100	24.1	1146	1	GNLJW2 HIV-1 retropepsin
36	1100	24.1	1146	1	GNLJEW HIV-1 retropepsin
37	1095	24.0	1101	1	B45390 HIV-1 retropepsin
38	1091	24.0	1101	1	GNLJVS HIV-1 retropepsin
39	1080.5	23.7	1145	1	GNLJEV HIV-1 retropepsin
40	1071	23.5	219	2	S32152 RNA-directed DNA p
41	1067	23.4	219	2	S32118 RNA-directed DNA p
42	1067	23.4	219	2	S32139 RNA-directed DNA p
43	1065	23.4	219	2	S32089 RNA-directed DNA p
44	1065	23.4	219	2	S32160 RNA-directed DNA p
45	1064	23.4	219	2	S32133 RNA-directed DNA p

#### ALIGNMENTS

##### RESULT 1

S54378

pol polyprotein - human immunodeficiency virus type 1 (fragment)

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004

C;Accession: S54378

R;Theodore, T.; Buckler-White, A.J.

submitted to the EMBL Data Library, July 1989

A;Reference number: S54377

A;Accession: S54378

A;Status: preliminary

A;Molecule type: genomic RNA

A;Residues: 1-1002 <THE>

A;Cross-references: UNIPROT:P12499; EMBL:M22639; NID:G329377; PIDN:AAA45366.1; PID:G3293

C;Superfamily: pol polyprotein

C;Keywords: polyprotein

Alignment Scores:			
Pred. No.:	3.6e-154	Length:	1002
Score:	3462.50	Matches:	651
Percent Similarity:	95.14%	Conservative:	34
Best Local Similarity:	90.42%	Mismatches:	28
Query Match:	76.02%	Indels:	7
DB:	2	Gaps:	4

US-09-610-313B-32 (1-2457) x S54378 (1-1002)

Qy	220	TTCTTCGCGAGGACCTGGCTTCCCGGAGGCGCGAGTTCCTCCCGAGCGAGCAG 279
Db	1	PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerSerGluGln 20
Qy	280	AACCGGCCAACGCCGCCAGCGAGCTGCGAGTGCCTGCGGC---GACAAACCCCGC 336
Db	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40
Qy	337	AGCGAGCGCGCGCGAGCGCGAGCGACCTGACCTTC-----CCCCAGATCACCTG 390
Db	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu 60



QY 391 TGGCAGCGCCCTTGTGAGCATCAAGTGGCGCCAGATCAAGAGAGCCCTGCTGGAC 450  
 Db 61 TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 80  
 QY 451 ACCGGCGCCAGCACACCGTGTGGAGGAGATGAGCCTGCCGGCAAGTGGAGGCCAAG 510  
 Db 81 ThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100  
 QY 511 ATGATCGGGGATCGGGCTTCATCAAGGTGGCCAGTAGCAGACAGATCCTGATCGAG 570  
 Db 101 MetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120  
 QY 571 ATCTCGGCACAAGSGCCATCGCACCGTGTGATCGGCCCCACCCCGTGAACATCATC 630  
 Db 121 IleCysGlyHisLeuAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140  
 QY 631 GGCCGCAACATCTGACCCAGCTGGGCTGCACCTGAACTTCCCATCAGCCCAATCGAG 690  
 Db 141 GlyArgAsnLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
 QY 691 ACCGTGCGGTGAAGTGAAGCCCGCATGAGCGGCCCAAGTGAAGCAGTGGCCCTG 750  
 Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180  
 QY 751 ACCGAGGAGAAGTCAAGCGCCTGACCGCATCTCGAGGAGATCGAGAGAGGGCAAG 810  
 Db 181 ThrGluGluLysIleLeuAlaLeuThrGluIleCysThrGluMetGluLysGluLys 200  
 QY 811 ATCAACAAGATCGGCCCGCCAGAACCCCTACAAACCCCGTGTCCCATCAAGAAGAAG 870  
 Db 201 IleSerArgValGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220  
 QY 871 GACAGACCAAGTGGCGCAAGCTGTGTGACTTCCGGAGCTGAACAAGGCCACCCAGGAC 930  
 Db 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240  
 QY 931 TTCTGGAGGTGACGTGGCATCCCAACCCCGCGCTCAAGAAGAAGAAGACGTG 990  
 Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysSerVal 260  
 QY 991 ACCGTGCTGGAGTGGCGACCCCTACTTTCAGCGTCCCTCGAGAGGACTTCCGCAAG 1050  
 Db 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLys 280  
 QY 1051 TACACCGCCTTCCATCCCAAGCATCAACAACGAGACCCCGGATCCGCTACCAAGTAC 1110  
 Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300  
 QY 1111 AACGTGCTGCCAGGCTGGAGGGCAGCCCGACGATCTTCCAGAGCATGCCCAAG 1170  
 Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 QY 1171 ATCTCGGAGCCCTTCGGCGCCGCAACCCCGAGATCGTGATCTACCAG-----GCCCC 1224  
 Db 321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleIleTyrGlnTyrMetAspAsp 340  
 QY 1225 CTGTACGTGGGAGCGACTGAGATCGGCCAGCACCGCGCAAGATCGAGAGTGGCG 1284  
 Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuArg 360  
 QY 1285 AAGCACCTGCTGGGTGGCTTCCACCAACCCCGCAAGACCAAGCAGGAGGCCCC 1344  
 Db 361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
 QY 1345 TTCTCTG-----CCCATCGAGTGCACCCCGCAAGTGCACCGTGCAGCCCATCGAGTGTG 1398  
 Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400  
 QY 1399 CCGGAGAGGAGGAGTGGACCGTGAACGACATCCAGAGCTGGTGGCGCAAGCTGAACTGG 1458  
 Db 401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

QY 1459 GCCAGCAGATCTACCCCGCATCAAGGTGGCGCCAGCTGTGCAAGCTGTGCGCGCGCC 1518  
 Db 421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440  
 QY 1519 AAGGCCCTGACGACATCGTCCCTGACCGAGAGGCGGAGCTGGAGCTGGCCGAGAAC 1578  
 Db 441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460  
 QY 1579 CGCGAGATCCTCGCGAGCGCGTGCACCGCTGTACTAGCACCCCGCAGCAAGACCTGGTG 1638  
 Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
 QY 1639 GCCAGATCCAGAAGCAGGCGCACACAGTGGAGCTTACCAGATCTACCAGAGCCCTTC 1698  
 Db 481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500  
 QY 1699 AAGAACCTGAAGCCGCAAGTACGCCAAGATGGCACCGCCCAACCAACGACGTGAAG 1758  
 Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520  
 QY 1759 CAGCTGACCGAGCGCTGCAGAGATCCCATGAGAGCATCGTGTGGGCAAGACC 1818  
 Db 521 GlnLeuAlaGluValValGlnLysIleSerThrGluSerIleValIleTrpGlyLysThr 540  
 QY 1819 CCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGGAGACCTGGTGACCGACTACTGG 1878  
 Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpValGluTyrTrp 560  
 QY 1879 CAGGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCCCTCGTGAAGCTGGTG 1938  
 Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
 QY 1939 TACAGCTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTAGCTGGAGCGCCGCC 1998  
 Db 581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 QY 1999 AACCOCGAGACCAAGATCGCAAGCCGCTACGTGACGACCGCGCGCGGCGAGAGATC 2058  
 Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
 QY 2059 GTGAGCTGACCGAGACCAACCAAGACCGAGTGCAGGCTGCAGGCTCCAGCTGGCCCTG 2118  
 Db 621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640  
 QY 2119 CAGGACAGCGCAGCGAGTGAACATCGTACCGACGACCGCTAGCGCTGGGCATCATC 2178  
 Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 QY 2179 CAGGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACACGATCATCGAGCGCTGATC 2238  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680  
 QY 2239 AAGAGGAGAAGTGTACTGAGCTGGTGGTGGCCGCCCAAGGGCATCGCGCGCAACGAG 2298  
 Db 681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700  
 QY 2299 CAGATCGACAGCTGGTGAAGCGGATCCGAGAGTGTCTTCTCGACCGGATCGAT 2358  
 Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 2

GNLJND

HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)

Nl:Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly

C:Species: human immunodeficiency virus type 1, HIV-1

A:Note: host Homo sapiens (man)

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004

C:Accession: JQ0067

R:Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.,

Gene 81, 275-284, 1989

A:Title: Nucleotide sequence of HIV-1-NDK: a highly cytopathic strain of the human immuno

A:Reference number: JQ0065; MUID:90034200; PMID:2806917

A:Accession: JQ0067

A: Molecule type: DNA  
 A: Residues: 1-1002 <SPI>  
 A: Cross-references: UNIPROT:P18802; GB:M27323; NID:G328154; PID:AAA44869.1; PID:G328158  
 C: Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
 C: Genetics:  
 A: Gene: pol  
 C: Superfamily: pol polyprotein  
 C: Keywords: ARDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle  
 F: 56-154/Product: retropepsin #status predicted <RTP>  
 F: 80/Active site: Asp (shared with dimeric partner) #status predicted

## Alignment Scores:

Pred. No.: 8,49e-154 Length: 1002  
 Score: 3454.50 Matches: 651  
 Percent Similarity: 95.14% Conservative: 34  
 Best Local Similarity: 90.42% Mismatches: 28  
 Query Match: 75.84% Indels: 7  
 DB: 1 Gaps: 4

US-09-610-313B-32 (1-2457) x GNLJND (1-1002)

QY 220 TTCTTCGCGAGGACCTGCGCTTCCCGAGGCGAGCCGCGAGTTCCCGAGCGAGCAG 279  
 DB 1 PhePheArgGluAspLeuAlaPheProGlnGlyysAlaGlyGluPheSerSerGluGln 20  
 QY 280 AACCGCGCAACAGCCACAGCGCGAGCTGCAGGTG---CGCGGCGCAACCCCGCGC 336  
 DB 21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu 40  
 QY 337 AGCGAGCGCGCGAGCGCGAGCGCGAGCCCGCTG-----ACTTCCCGCAGATCACCGTG 390  
 DB 41 SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnLeuThrLeu 60  
 QY 391 TGGCAGCGCCCTCGTGAGCATCAAGGTGGCGCGCAGATCAAGGAGCGCCCTCGTGCAC 450  
 DB 61 TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 80  
 QY 451 ACCGCGCGCAGCACCGCTGCTGGAGGAGATGAGCTCGCCCGCGCAAGTGGAGGCCCAAG 510  
 DB 81 ThrGlyAlaAspAspThrValLeuGluGluLeuAsnLeuProGlyLysTrpLysProLys 100  
 QY 511 ATGATCGCGCGCATCGCGGCTTATCATGAGTGGCGCGAGTACGACGAGATCCTGATCGAG 570  
 DB 101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnLeuLeuIleGlu 120  
 QY 571 ATCTGCGCAAGAGGCCATCGGCACCGTGTGATCGSCCCACCCCGCTGAACATCATC 630  
 DB 121 IleCysGlyTyrlsAlaMetGlyThrValLeuValGlyProThrProValAsnIlelle 140  
 QY 631 GGCGCGCAACATGTCAGCCAGCTGGCTGCAACCTGAACTTCCCGCATCAGCCCGCATCGAG 690  
 DB 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
 QY 691 ACCGTGCGCGTGAAGTGAAGCCCGGATGACCGCGCGCCCGCGCGCGCGCGCGCGCGCTG 750  
 DB 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180  
 QY 751 ACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGCGAGGAGATGGAGAGGGCGCAAG 810  
 DB 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200  
 QY 811 ATCAAGAGATCGGCGCGCGAGAACCCCTTACAAACACCCCGCTGTTCGCGCATCAAGAGAAG 870  
 DB 201 IleSerArgIleGlyProGluAsnProTyrlsAsnThrProIlePheAlaIleLysLysLys 220  
 QY 871 GACAGCAACAGTGGCGCAGCTGGTGAATTCGCGAGCTGAACAGCGCACCCAGCAG 930  
 DB 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240  
 QY 931 TTCTGGGAGGTGAGCTGGGATCCCCACCCCGCGCTGAGAGAGAGAGAGAGCGTGTG 990  
 DB 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260

QY 991 ACCGTGTGGACGTGGGCGACGCTTACTTACGCGTGGCCCTGGACGAGACTTCGCGAAG 1050  
 DB 261 ThrValLeuAspValGlyAspAlaTyrlsPheSerValProLeuAspGluAspPheArgLys 280  
 QY 1051 TACACCGCTTACCATCCCGCAGCATCAACACGAGACCCCGCGCATCGCTACAGTAC 1110  
 DB 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrlsGlnTyr 300  
 QY 1111 AACGTGTGCTGCCCGCAGGCTGGAAGCGCAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAG 1170  
 DB 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 QY 1171 ATCTGAGAGCCTTCCGCGCGCCCAACCCCGAGATCGTGTATCTACAG-----GCCGCC 1224  
 DB 321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrlsGlnTyrMetAspAsp 340  
 QY 1225 CTGTACCTGGCGCAGCGACCTGGAGATCGCGCAGCAGCCCGCGCAGATCGAGAGCTGGCG 1284  
 DB 341 LeuTyrlsValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360  
 QY 1285 AAGCACCTGTGCTGGCGGCTTCCACCAACCCCGCAGAACAGACCAAGAGAGCCCGCC 1344  
 DB 361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
 QY 1345 TTCTGT-----CCCATCGAGCTGCACCCCGCAGTGGACCGTGCAGCCCATCGAGCTG 1398  
 DB 381 PheLeuTrpMetGlyTyrlsGluLeuHisProAspLysTrpThrValGlnProIleAsnLeu 400  
 QY 1399 CCCGAGAGGAGAGCTGGACCGTGAAACGACATCCAGAGCTGTGGGCGAGCTGACTGG 1458  
 DB 401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
 QY 1459 GCCAGCAGATCTACCCCGCGCATCAAGTGGCGCAGCTGTCAAGCTGCTCGCGCGGCC 1518  
 DB 421 AlaSerGlnIleTyrlsAlaGlyIleLysValLysGlnLeuLysLysLeuLeuArgGlyThr 440  
 QY 1519 AAGCCTGTACCGACATCGTGTGCTGCGCCCTGACCGAGAGCGCGAGCTGGAGCTGGCAGAAC 1578  
 DB 441 LysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuLeuAlaGluAsn 460  
 QY 1579 CGCGAGATCTGCGCGAGCGCGTGCACGGGTGTACTACGACCCCGCAGCAGCAGCTGTG 1638  
 DB 461 ArgGluIleLeuLysGluProValHisGlyValTyrlsAspProSerLysAspLeuIle 480  
 QY 1639 GCCGAGATCCAGAACGAGGCGCACGACGAGTGGACCTTACAGATCTTACAGAGCGCTTC 1698  
 DB 481 AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrlsGlnIleTyrlsGlnLysProPhe 500  
 QY 1699 AAGAACCTGAAGACCGCAAGTACGCCAAGATGCGCACCGCCCGCCACACCAACGAGCTGAAG 1758  
 DB 501 LysAsnLeuLysThrGlyLysTyrlsAlaArgThrArgGlyAlaHisThrAsnAspValLys 520  
 QY 1759 CAGCTGACGAGCGCGTGCAGAGATCGCCATGAGAGCATCGTGTGATCTGGGCGCAGAC 1818  
 DB 521 GlnLeuThrGluAlaValGlnLysIleAlaThrLysSerIleValIleTrpGlyLysThr 540  
 QY 1819 CCAGATGCTCGCTGCGCCCATCCAGAGGAGACCTGGGAGACCTGTGTGACGACTACTTGG 1878  
 DB 541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrlsTrp 560  
 QY 1879 CAGGCCACCTGGATCCCGCGAGTGGAGTTCGTGAACACCCCGCTGTGGTGAAGCTGTGG 1938  
 DB 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
 QY 1939 TACAGCTGAGAGAGGAGCGCATCATCGGCGCGAGACCTTCTAGCTGGAGCGCGCGCGCC 1998  
 DB 581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrlsValAspGlyAlaAla 600  
 QY 1999 AACCGCAGAGACCAAGATCGCGCAGCGGCTTACGTGACCGACCGCGCGCGCGCAGCAAGATC 2058  
 DB 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrlsValThrAspArgGlyLysGlnLysVal 620  
 QY 2059 GTGAGCCTGACCGAGAGACCAACCAAGAGACCGAGCTGCAGCGCCATCCAGCTGGCCCTG 2118





```
Db 281 AlaTyrPheSerValProLeuAspGluAspPheAsgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACAGAGACCCCGGATCGCTACCAAGTACAAAGTCTGCTGCCAGCGGTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGCGACCCAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGCCCTTCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerMetThrLysIleLeuGluProPheLysLys 340
QY 1192 CGCAACCCCGAGATCGTATCTACACG-----GCCCCCTGTACGTGGCGAGCACCTG 1245
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
QY 1246 GAGATCGGCGACACCGGCCCAAGATCGAGAGCTGGCAGACCTGCTGGCTGGGGC 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
QY 1306 TTCACCCCGCCGACAAGACACAGAGAGAGCCCTTCCTG-----CCCATCGAG 1359
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1360 CTGCACCCCGACAAGTGGACCTGCGGCCATFCGAGCTGCCGAGAGAGAGAGCTGGACC 1419
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1420 GTGAACGACATCCAGAGCTGTGGGCAAGCTGAAGCTGGCGCAGCGACGATCTACCCGCG 1479
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
QY 1480 ATCAGGTGGCGCAGCTGCAAGCTGTGGCGGCGCCAGGCCCTGACCGACATCTGTG 1539
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1540 CCCCTGACCGAGGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGGACGCC 1599
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLysGluPro 480
QY 1600 GTGCGCGGCTGTACTAGACCCCGAGAGAGCTGTGGCGGCGAGATCCAGAGCAGGCG 1659
Db 481 ValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1660 CACGACCACTGACCTACAGATCTACAGAGAGCCCTTCAAGAACTCAAGACCGGCAAG 1719
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1720 TACGCCAAGATCGGCACCGCCACACCAACGACGTGAAGCAGCTGACCGAGCGCTGCAG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1780 AAGATCGCCATCGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTGCGCTGCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTyrGlyLysThrProLysPheLysLeuProIle 560
QY 1840 CAGAGGAGACTGGGAGACCTGTGGACCGACTACTGGCGGCCACCTGGATCCCGCAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnIleTyrPglAlaThrTrpIleProGlu 580
QY 1900 TGGGAGTTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTGACAGCTGGAGAGAGAGCC 1959
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuLysGluPro 600
QY 1960 ATCATCGGCGCGAGACCTTCTACTGTGACGCGCGCGCAACCGCGAGACCAAGATCGGC 2019
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCGCGCTAGTCAACGACCGGCGCGGAGAGATCGTACGCTGACCGAGACCAACC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACCAAGAACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGGTG 2139
```

```
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2140 AACATCTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGACGACAAGAGC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGACGAGCTGGTGAACACGAGATCATCCAGCAGCTGTCAAGAGGAGAGAGGTGTACCTG 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTGCCCGCCCAACAGGGCATCGCGCAACGAGCAGATCGACAAGTGTGTAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGlnGlnValAspLysLeuValSer 720
QY 2320 AAGGCGATCCCGACAGGTCTCTTCGACGCGCATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 5
T09440
pol polyprotein - human immunodeficiency virus type 1 (strain JRF1) (fragment)
C:Species: human immunodeficiency virus type 1, HIV-1
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T09440
R:Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.-S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A:Reference number: Z16673
A:Accession: T09440
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1003 <PAN>
A:Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 1,01e-151 Length: 1003
Score: 3410.00 Matches: 642
Percent Similarity: 94.45% Conservative: 39
Best Local Similarity: 89.04% Mismatches: 32
Query Match: 74.86% Indels: 8
DB: 2 Gaps: 4

US-09-610-313B-32 (1-2457) x T09440 (1-1003)
QY 220 TTCTTCCGCGAGGACCTGGCTTCCCGCAGGCGAAGCCCGCGAGTTCCCGACGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCAACAGCCCGCCAGCCGCGAGCTGCAGGTG-----CGCGGCGACACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTrpGlyArgAspSerAsnSer 40
QY 334 CGCAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlnValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCTCTGTGAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCGCGCGCGCGCGCGCGCGCG 567
Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeu 120
QY 568 GAGATCTCCGCGCAAGAAGCCCATCGCACCTGTGTGATCGCGCGCGCGCGCGCGCGCGCG 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
```





Alignment Scores:

Pred. No.: 1,39e-151 Length: 1003  
 Score: 3407.00 Matches: 643  
 Percent Similarity: 94.31% Conservative: 37  
 Best Local Similarity: 89.18% Mismatches: 33  
 Query Match: 74.80% Indels: 8  
 DB: 1 Gaps: 4

US-09-610-313B-32 (1-2457) x B44001 (1-1003)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCAAGGCGCGAGTTCCTCCGCGAGCAG 279  
 Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20  
 QY 280 AACCGCGCCACAGCCCAACAGCCGCGAGCTGCGAGTG- - - - -CGCGGCGACAAACCC 333  
 Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgAspAsnAsnSer 40  
 QY 334 CGCAGCGAGGCGCGCGCGAGCGCGAGGCGACCTCG- - - - -AAGTTCCTCCCGAGATCAC 387  
 Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60  
 QY 388 CTGTGCGCAGCGCCCTGTGTAGCATCAAGTGGCGCGCGAGTCAAGAGGCGCCCTGTCTG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCGCGCGCGAGCACACCGTGTGTGAGGAGATGAGCTGCCCGGCAAGTGGAGGCC 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGAGTCAAGAGTGTGATC 567  
 Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgIleLysValSerPheProGlnIleThr 120  
 QY 568 GAGATCTGCGCGCAAGAAGCCATCGGCACCTGTGTGATCGCGCGCCCAACCTGTGAACATC 627  
 Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGCGCGCAACATGTCGACCCAGCTGAAGCGCGCATGGAGCGCGCCCAAGTGAAGTGGCGCC 747  
 Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACGCTGCGCGTGAAGCTGAAGCGCGCATGGAGCGCGCCCAAGTGAAGTGGCGCC 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGGAGGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAGGAGGCG 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCACCAAGATCGCGCGCGAGAACCCCTCAACACCCCGTGTTCGCCATCAAGAG 867  
 Db 201 LysIleSerLysIleGlyProGluAsnProTrpAsnThrProValPheAlaIleLysLys 220  
 QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGTGGAGCTTCGCGAGCTCAACAAAGCGCACCCAG 927  
 Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
 QY 928 GACTCTCTGGAGGTGCGTGGGCGATCCCGACCCCGCGCGCTGAAAGAAAGAGAGC 987  
 Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 989 GTGACCGTCTGGAAGTGGCGCGAGCTTCTTACGGTGTCCCTTGAGAGGAGTTCGCG 1047  
 Db 261 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuHisGluAspPheArg 280  
 QY 1048 AAGTACACCGCTTCACCATCCCGAGCATCAACAGAGACCCCGCGCATCGCTTACCGAG 1107  
 Db 281 LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyThrArgTrpGln 300  
 QY 1108 TACAACGTGCTCGCCCGAGGCTGGAGGGCGAGCCCGAGCATCTTCAGAGCATGATGCC 1167  
 Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

QY 1168 AAGATCTCTGGAGCCCTTCGCGCGCGCAACCCCGAGATCGTGATCTACCAG- - - - -GCC 1221  
 Db 321 ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTrpGlnTrpMetAsp 340  
 QY 1222 CCCCTGTACGTGGCGAGCAGCTGGAGATCGCGCGAGCACCGCGCGCAAGATCGAGGAGCTG 1281  
 Db 341 AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 QY 1282 CGCAAGCACCTGTCTGCGCTGGGGCTTACCAACCCCGCAAGAGACCAAGAGAGGCC 1341  
 Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCCTG- - - - -CCCATCGAGTGCACCCCGACAAAGTGCAGCGTGCAGCCCATCCAG 1395  
 Db 391 ProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
 QY 1396 CTGCCCGAAGAGGAGCTGGACCGTGAACACATCCAGAGCTGTGGGCAAGCTGAAC 1455  
 Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
 QY 1456 TGGCGCAGCAGATCTACCCCGGATCAAGGTGGCGCGAGCTGTGCAAGCTGTGCGCGCG 1515  
 Db 421 TrpAlaSerGlnIleTrpAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
 QY 1516 GCCAAGGCCCTGACGACATCGTCCCTGACCGAGGAGCGCGAGTGCAGTGCAGCCGAG 1575  
 Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460  
 QY 1576 AACCGCAGATCTTCGCGAGCCCGTGCACCGCGCTGTACTACGACCCCGAGCAAGACCTG 1635  
 Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTrpAspProSerLysAspLeu 480  
 QY 1636 GTGGCGCAGATCCAGAGCAGCGCCACACAGTGGAGCTTACCATCTACAGAGGCC 1695  
 Db 481 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTrpGlnIleTrpGlnGluPro 500  
 QY 1696 TTCAAGAACCTGAAGACCGCGAGTACGCCAAGATCGCCACCGCCACCAACCGACCTG 1755  
 Db 501 PheLysAsnLeuLysThrGlyLysTrpAlaArgThrArgGlyAlaHisThrAsnAspVal 520  
 QY 1756 AAGCAGCTGACCGAGCGCGTGCAGAAAGATCGCCATCGAGAGCATCTGTGGGCAAG 1815  
 Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
 QY 1816 ACCCCAAAGTTCGCGCTCCCATCCAGAAAGAGACCTGGAGACCTGTGAGCCACTAC 1875  
 Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTrp 560  
 QY 1876 TGGCAGGCCACTCGATCCCGAGTGGGAGTTTCGTGAACACCCCGCTGTGTGAAGCTG 1935  
 Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
 QY 1936 TGTGTACAGCTGGAGAGGCCCATCATCTCGCGCGCGAGACCTTCTACGTGAGCGCGCC 1995  
 Db 581 TrpTrpGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTrpValAspGlyAla 600  
 QY 1996 GCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGCGCGCGAGAG 2055  
 Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTrpValThrAsnLysGlyArgGlnLys 620  
 QY 2056 ATCTGTAGCCCTGACCGAGCACCAACCGAGAGCCGAGCTGCAGCGCTCCAGCTGGCC 2115  
 Db 621 ValValSerLeuThrAspThrAsnGlnLysThrGluLeuGlnAlaIleTrpLeuAla 640  
 QY 2116 CTGACGAGCAGCGCGAGGTGAACATCTGTGACCGCAGCAGCAGTACGCTGGGCATC 2175  
 Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTrpAlaLeuGlyIle 660  
 QY 2176 ATCCAGGCCCGCCGACAGAGCGAGCGAGCTGGTGAACAGCATCATTCGAGGAGCTG 2235  
 Db 661 IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleGluGlnLeu 680



```
QY 1696 TTCAAGAACCTGACGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGACGTG 1755
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1756 AAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGCGAGAGCATCGTGATCTGGGGCAAG 1815
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1816 ACCCCAGTTCCGCTCGCCATCCAGAGAGAGACTGGAGACCTGGAGACCGACTAC 1875
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaIleTrpMetGluTyr 560
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1876 TGGCAGGCGCACCTGATCCCGAGTGGAGTTCGTAACACCCCGCTGGTGAAGCTG 1935
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1936 TGTGTACCACTGGAGAGAGCCCATCATCGCGCGCGAGACCTTCTAGCTGGACGCGCC 1995
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 581 TrpTyrGlnLeuLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1996 GCCAACCGGAGACCAAGATCGGCAAGCGCGGCTACGTACCGACCGGGGCGCGAGAAG 2055
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2056 ATCGTGAGCTGACCGAGACACCAACAGAGAGACCGAGCTGACGCGCATCCAGCTGGCC 2115
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2116 CTGACGAGCAGCGGCGAGGTGAACATCGTGACCGACGACGACGCGCTGGGCGATC 2175
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2176 ATCCAGGCGCCAGCCGACGAGAGAGAGCGAGCTGGTGAACCATCATCGAGCAGCTG 2235
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2236 ATCAAGAGGAGAGAGTGTACCTGAGTGGTGGCGCCGCCAACAGCGGATCGCGGCAAC 2295
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2296 GAGCAGATCGACAAGCTGGTGAGCAAGGCGCATCCGCAAGGTGCTGTTCTGTGAGCGCATC 2355
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2356 GAT 2358
Db |||
QY 721 Asp 721
Db |||

RESULT 8
GNVWL
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LV)
N:Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
A:Accession: A03967
R:Muessing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.
Nucleur 313, 450-456, 1985
A:Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi
A:Reference number: A93355; MUID:85111157; PMID:2982104
A:Accession: A03967
A:Molecule type: DNA
A:Residues: 1-1012 <MUE>
A:Cross-references: UNIPROT:P03368
C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein
C:Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle
F:66-164/Product: retropepsin #status predicted <ATP>
F:90/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:
Pred. No.: 4.53e-151 Length: 1012
```

```
Score: 3396.00 Matches: 643
Percent Similarity: 93.12% Conservative: 34
Best Local Similarity: 88.45% Mismatches: 30
Query Match: 74.56% Indels: 20
DB: 1 Gaps: 5
US-09-610-313B-32 (1-2457) x GNVWL (1-1012)

QY 238 GCCTTCCCGCAGGCGAAGCGCGCGAGTTC----- 267
Db |||||||||||||||||||||||||||||||||||||||||||
QY 4 AlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerPro 23
Db |||||||||||||||||||||||||||||||||||||||||||
QY 268 -----CCACGAGCAGAACCGCGCCAAACAGCCCGAGCTGCAGGTG--- 318
Db |||||||||||||||||||||||||||||||||||||||||||
QY 24 ThrIleSerSerGluGlnThrArgAlaAsnSerProThrArgArgGluLeuValTrp 43
Db |||||||||||||||||||||||||||||||||||||||||||
QY 319 ---CGCGCGCAGAACCCCGCAGGAGCGCGCGCGAGCGCCAGCCCGT----- 369
Db |||||||||||||||||||||||||||||||||||||||||||
QY 44 GlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPhe 63
Db |||||||||||||||||||||||||||||||||||||||||||
QY 370 AACTTCCCGCAGATCACCTGTGTGGCAGCGCGCGCTGTGAGCATCAAGTGGCGCGCCAG 429
Db |||||||||||||||||||||||||||||||||||||||||||
QY 64 AsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGlyGln 83
Db |||||||||||||||||||||||||||||||||||||||||||
QY 430 ATCAAGGAGGCGCTGTGTGACACCGCGCGCGCGAGCACCGTGTGGAGAGATGAGCCTG 489
Db |||||||||||||||||||||||||||||||||||||||||||
QY 84 LeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluMetSerLeu 103
Db |||||||||||||||||||||||||||||||||||||||||||
QY 490 CCGCGCAATGGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCCAG 549
Db |||||||||||||||||||||||||||||||||||||||||||
QY 104 ProGlyArgTyrLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGln 123
Db |||||||||||||||||||||||||||||||||||||||||||
QY 550 TAGCAGCAGATCTGTGATCGAGATCTGCGCGCAAGAGCGCATCGCGCTGTGATCGCGC 609
Db |||||||||||||||||||||||||||||||||||||||||||
QY 124 TyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIleGlyThrValLeuValGly 143
Db |||||||||||||||||||||||||||||||||||||||||||
QY 610 CCCACCCCGGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGACCCCTGAGAAC 669
Db |||||||||||||||||||||||||||||||||||||||||||
QY 144 ProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsn 163
Db |||||||||||||||||||||||||||||||||||||||||||
QY 670 TTCCCATCAGCCCATCGAGACCGTGCCTGTAAGTGAAGCCCGGATGAGACCGCGCC 729
Db |||||||||||||||||||||||||||||||||||||||||||
QY 164 PheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyPro 183
Db |||||||||||||||||||||||||||||||||||||||||||
QY 730 AAGTGAAGCAGTGGCGCGCTGACCGAGGAGATCAAGGCGCTGACCGCATCTGCGAG 789
Db |||||||||||||||||||||||||||||||||||||||||||
QY 184 LysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThr 203
Db |||||||||||||||||||||||||||||||||||||||||||
QY 790 GAGATGGAAGAGGAGCGCAAGATCACCAAGATCGCGCGCGAGAACCCCTTACACACCCCC 849
Db |||||||||||||||||||||||||||||||||||||||||||
QY 204 GluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrPro 223
Db |||||||||||||||||||||||||||||||||||||||||||
QY 850 GTGTTTCGCATCAAGAAAGAGACAGACCAAGTGGCGGCAAGCTGGTGGACTTCCGCGAG 909
Db |||||||||||||||||||||||||||||||||||||||||||
QY 224 ValPheAlaIleLysLysLysAspSerThrLysTyrArgLysLeuValAspPheArgGlu 243
Db |||||||||||||||||||||||||||||||||||||||||||
QY 910 CTGAACAGCGCACCCAGAGATCTTCTGGAGGTGCGAGTGGGCATCCCGCCCGCGCGC 969
Db |||||||||||||||||||||||||||||||||||||||||||
QY 244 LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGly 263
Db |||||||||||||||||||||||||||||||||||||||||||
QY 970 CTGAAGAAGAAGAGCGTGTACCGTGTGACGTGGCGCGCGCTTCTTACGCGTGGCC 1029
Db |||||||||||||||||||||||||||||||||||||||||||
QY 264 LeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValPro 283
Db |||||||||||||||||||||||||||||||||||||||||||
QY 1030 CTGAGCAGGAGTTCGCAAGTACACCGCTTACCATCCCGCAGCATCAACACGAGACC 1089
Db |||||||||||||||||||||||||||||||||||||||||||
QY 284 LeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThr 303
Db |||||||||||||||||||||||||||||||||||||||||||
QY 1090 CCGCGCATCCGTACAGTACACCTGTGCGCGCGAGGTGCGAGGCGAGCCCGCGAGCATC 1149
Db |||||||||||||||||||||||||||||||||||||||||||
QY 304 ProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIle 323
Db |||||||||||||||||||||||||||||||||||||||||||
QY 1150 TTCCAGAGCAGCATGACCAAGATCTTCCGCGCGCGCGCTTCCGCGCGCGCAACCCCGAGATCGTG 1209
Db |||||||||||||||||||||||||||||||||||||||||||
```

Db 324 PheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343  
QY 1210 ATCTACCCAG-----GCCCCCTGTGTACCTGGCGACGACCTGGAGATCGGCACGACCGC 1263  
Db 344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363  
QY 1264 GCCAAGATCGAGAGCTGGCAAGACCTGTGCGCTGGGGTTCCACACCCCGACAAAG 1323  
Db 364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTTPGlyLeuThrThrProAspLys 383  
QY 1324 AAGCACCAGAGAGCCCTCTCTG-----CCATCGAGCTGCACCCCGACAGTGG 1377  
Db 384 LysHisGlnLysGluProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyr 403  
QY 1378 ACCGTGACGCCATCGAGCTGCCGAGAGAGCTGCACCGTGAACGACATCCAGAAAG 1437  
Db 404 ThrValGlnProIleValLeuProGluLysAspSerTyrThrValAsnAspIleGlnLys 423  
QY 1438 CTGTGTGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCAGCTG 1497  
Db 424 LeuValGlyLysLeuAsnTyrAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443  
QY 1498 TGCAAGCTGCTGGCGGCGCCAAAGCCCTGACGACATCTGCTGCCCTACCGAGAGGCC 1557  
Db 444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463  
QY 1558 GAGCTGGAGCTGGCGGAGAACCGGAGATCTTGGCGAGCCCGTGCAGCGGTGTACTAC 1617  
Db 464 GluLeuGluLeuAlaGluAsnArgGluLeuLysGluProValHisGlyValTyrTyr 483  
QY 1618 GACCCCAAGAGACCTGTGTGGCGAGATCCAGAACGAGGCGCCACGACCTGACCTAC 1677  
Db 484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTyrTyrTyr 503  
QY 1678 CAGATCTACGAGGACCTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATCGGCAC 1737  
Db 504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523  
QY 1738 GCCCACACCAACGAGCTGAACGAGCTGACCGAGCGCGTGCAGAGATCGCCATCGAGAGC 1797  
Db 524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543  
QY 1798 ATCTGTATCTGGGCAAGACCCCAAGTTCGCTGCCCTGCCATCCAGAAAGAGACCTGGAG 1857  
Db 544 IleValIleTyrGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTyrGlu 563  
QY 1858 ACCTGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTCGTGAACAC 1917  
Db 564 ThrTyrTyrThrGluTyrTyrGlnAlaThrTyrIleProGluTyrGluPheValAsnThr 583  
QY 1918 CCCCCCTGTGTGAAGCTGTGTGTACCGAGTGGAGAGGAGCCCATCATCGGCGCGAGACC 1977  
Db 584 ProProLeuValLysLeuTyrTyrGlnLeuGlnLysGluProIleValGlyValGluThr 603  
QY 1978 TTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTGACC 2037  
Db 604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGlyLysAlaGlyTyrValThr 623  
QY 2038 GACCGGCGCCCGCAGAGATCGTGAGCTGACCGAGACCAACCAACCAAGAACCGAGCTG 2097  
Db 624 AsnLysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643  
QY 2098 CAGCGCATCCAGCTGGCCCTCAGCAGCGCGGAGCGGTGACATCGTACCGCAGC 2157  
Db 644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSer 663  
QY 2158 CAGTACGCGCTGGGCATCATCAGCGCCAGCCCAACAGACGAGAGCGAGCTGTGTGAAC 2217  
Db 664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsn 683  
QY 2218 CAGATCATCGAGCTGTATCAAGAGAGAGTGTACTCTGAGCTGGTCCCGCCAC 2277

Db 684 GlnIleIleGluGlnLeuIleLysGluLysValTyrLeuAlaTyrValProAlaHis 703  
QY 2278 AAGGCGATCGCGCGCAACGACGATCGACAGCTGTGTGAGCAAGGCGATCCGCAAGGTG 2337  
Db 704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723  
QY 2338 CTGTTCTCTGGACCGCATCGAT 2358  
Db 724 LeuPheLeuAspGlyIleAsp 730  
RESULT 9  
GNLSI  
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz  
N:Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
C:Species: simian immunodeficiency virus SIVcpz  
A:Note: host Pan troglodytes (chimpanzee)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C:Accession: S09984  
R:Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.  
Nature 345, 356-359, 1990  
A:Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.  
A:Reference number: S09983; MUID:90259077; PMID:2188136  
A:Accession: S09984  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1027 <HUE>  
A:Cross-references: UNIPROT:P17283; EMBL:X52154  
C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
C:Genetics:  
A:Gene: pol  
C:Superfamily: pol polypeptide  
C:Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo  
F:81-180/Product: retropepsin #status predicted <RTP>  
F:105/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
Pred. No.: 1,93e-140 Length: 1027  
Score: 3168.00 Matches: 591  
Percent Similarity: 88.38% Conservative: 63  
Best Local Similarity: 79.86% Mismatches: 79  
Query Match: 69.55% Indels: 7  
DB: 1 Gaps: 3

US-09-610-313B-32 (1-2457) x GNLSI (1-1027)

QY 159 CTGCTGGAAGTGGCGGCAAGGAGGCGCCACGATGAAGGACTGCACCGAGCGCCAGGCCAA 218  
Db 7 LeuLeuAlaValTyrAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26  
QY 219 CTTCTTCCTCGGAGACCTGCGCTTCCCGCGGCAAGCGCGAGTTCCTCCAGCGAGCA 278  
Db 27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGlu 46  
QY 279 GAACCGCGCCAAAGCCCGACCGCGAGCTGCAGGTGCGCGC-----GACAAACC 332  
Db 46 nAsnArgThrAsnGlyProThrAspArgGluLeuTyrValProGlyGlyArgGluLys 66  
QY 333 CCGCAGCGAGCG 392  
Db 66 oGlyGluGluArgGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86  
QY 393 GCACGCGCCCTGTGTGAGCATCAAGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452  
Db 86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106  
QY 453 CGCGCGCGCGACACCTGTGTGGAGAGATGAGCTGCCCGCGAGTGGAGAGCCCAAGAT 512  
Db 106 xGlyAlaAspThrValIleGluArgIleGlnLeuGlnGlyLeuTyrLysProLysMe 126  
QY 513 GATCGCGCGCATCGCGCGCTTCATCAAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 572  
Db 126 tileGlyGlyIleGlyPheIleLysValLysGlnPheAspAsnValHisIleGluI 146



Qy	481	ATGAGCTTCGCGGCAAGTGAAGCCCAAGATGATCGGCGCGCATCGCGCGCTTCATCAAG	540
Db	1	MetAsnLeuProGlyAArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys	20
Qy	541	GTGCGCCAGTACGACACAGATCCTCATCGAGATCTGCGCAAGAGGCCATCGGACCGTGC	600
Db	21	ValLysGlnTyrAspGlnIleAlaIleGluIleCysGlyHisLysAlaIleGlyThrVal	40
Qy	601	CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATCTGAGCCACCGCTGGCTGC	660
Db	41	LeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCys	60
Qy	661	ACCCTGAACCTCCCATCAGCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGCATG	720
Db	61	ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet	80
Qy	721	GACGCGCCCAAGTGAAGCAGTGCCTGTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780
Db	81	AspGlyProLysValLysGlnTrpProLeuThrGluGlnLysIleLysAlaLeuIleGlu	100
Qy	781	ATCTGCGGAGAGATGGAAGAGGCGCAAGATCACCAAGATCGCGCCCGCAGAACCCCTAC	840
Db	101	IleCysThrGluMetGluLysGluLysIleSerLysIleGlyProGluAsnProTyr	120
Qy	841	AACACCCCGTGTTCGCCATCAAGAGAAGACAGACCAAGTGGCGGAAGCTGGTGGAC	900
Db	121	AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp	140
Qy	901	TTCGCGAGCTGAACACGCGACCCAGGACTTCTGGGAGGTGCGAGCTGGCATCCCCCAC	960
Db	141	PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis	160
Qy	961	CCCGCCGCTGAAGAGAAGAGAGCGGTGACCGTGTGACGTGGAGTGGCGCAGCCCTACTTC	1020
Db	161	ProGlyGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe	180
Qy	1021	AGCTGTCCTTGAACGAGGACTTCGCGAAGTACACCCGCTTCACCATCCCGAGCATCAAC	1080
Db	181	SerValProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsn	200
Qy	1081	AACGAGACCCCGCATCCGCTACCACTACAACTGTGTCGCCAGGCGCTGGAGGGCAGC	1140
Db	201	AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer	220
Qy	1141	CCAGCATCTTCCAGAGCAGCATCACCAAGATCTCGAGGCCTTCGCGCGCCGCAACCCC	1200
Db	221	ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro	240
Qy	1201	GAGATCGTGAATCTACCAG-----GCCGCCCTGTACTGTGGCAGCAGCTGGAGATCGGC	1254
Db	241	GluIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly	260
Qy	1255	CAGCACCGCCCAAGATCGAGGAGCTGGCAGACCTGTGCTGCTGGCTGGGCTTCACCAAC	1314
Db	261	GlnHisArgThrLysIleGluGluLeuArgArgHisLeuLeuArgTrpGlyPheThrThr	280
Qy	1315	CCCGACAAGAAGCACCAAGAGGCCCCCTCTCTCTGCCCATC-----GAGCTGCACCCC	1368
Db	281	ProAspLysLysHisGlnLysGluProProPheLeuTrpIleGlyTyrGluLeuHisPro	300
Qy	1369	GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGACCGTGAACGAC	1428
Db	301	AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp	320
Qy	1429	ATCCAGAAGCTGGTGGGCAAGCTGAATGGGCGACGACAGATCTACCCCGGCGCATCAAGTG	1488
Db	321	IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal	340
Qy	1489	CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGTACCGACATCTGTGCCCTCACC	1548
Db	341	ArgGlnLeuCysArgTrpLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr	360
Qy	1549	GAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGCGAGCCGCTGCACGC	1608

Db	361	LySGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysThrProValHisGly	380
QY	1609	GTGTFACCTACGACCCACCAAGGACCTCGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAG	1668
Db	381	VallTyTrpAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyLeuGlyGln	400
QY	1669	TGGACCTACCAAGATCTACCAAGGACCCCTTCAAGAACCTGAAGACCGCAAGTAGCCCAAG	1728
Db	401	TrpThrTyrglnIleTyrglnGluProPheLysAsnLeuLysThrGlyLysTyralaLys	420
QY	1729	ATGGGCACCGCCACACCAACGAGCTGAACGAGCTGACCGAGGCGGTGCAGAGAGATCGCC	1788
Db	421	MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla	440
QY	1789	ATGAGAGAGCATCGTATCTGGGCAACACCCCAAGTTCGCTGCCCTCCATCCAGAAGGAG	1848
Db	441	ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu	460
QY	1849	ACCTGGGAGAGACTCGTGGACCGACTACTGGCAGCGCACCTCGATGCCCGAGTGGAGTTC	1908
Db	461	ThrTrpGluAlaTrpTrpMetGluTyrrTrpGlnAlaThrTrpIleProGluTrpGluPhe	480
QY	1909	GTGAACACCCCCCTGGTGAAGCTGTGTATCCAGCTGGAGAGAGCCCATCATCGCC	1968
Db	481	ValAsnThrProProLeuValLysLeuTrpTyrrGlnLeuGluLysGluProIleValGly	500
QY	1969	GCCGAGACCTTCTACGTGGAGCGCGCCGACCAACCGGAGACCAAGATCGGCACAGGCCGC	2028
Db	501	AlaGluThrPheTyrrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly	520
QY	2029	TACGTGACCAACCGGGCCGCGAAGATCGTGAAGCTGACCGAGACCAACCAACAGAAAG	2088
Db	521	TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrThrAsnGlnLys	540
QY	2089	ACCGAGCTGAGGCCATCCAGCTGGCCCTCGAGACACGCGCAGCGAGGTGAACATCGTG	2148
Db	541	ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal	560
QY	2149	ACCGACACCGATACGCGCCCTGGGCATCATCCAGGCCCGCCGACAGAGCGAGAGCGAG	2208
Db	561	ThrAspSerGlnTyrrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu	580
QY	2209	CTGTGTAAACAGATCATCTGAGCAGCTGATCAAAGAGGAGAAGTGTACCTGAGCTGGGTG	2268
Db	581	LeuValSerGlnIleIleGluHisLeuIleLysLysGlnLysValTyrrLeuAlaTrpVal	600
QY	2269	CCCGCCCAACAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGTGACCAAGGCGCATC	2328
Db	601	ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIle	620
QY	2329	CGCAAGTGCTGTCTCGAGCGGCATCGAT	2358
Db	621	ArgLysValLeuPheLeuAspGlyIleAsp	630

## RESULT 11

rs2501 11  
 T01668  
 pol polyprotein - human immunodeficiency virus type 1  
 C;Species: human immunodeficiency virus type 1, HIV-1  
 C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text  
 C;Accession: T01668  
 R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.  
 Cell 46, 63-74, 1986  
 A;Title: Genetic variability of the AIDS virus: Nucleotid  
 A;Reference Number: Z14389; MUID:86245056; PMID:2424612  
 A;Accession: T01668  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-902 <all>  
 A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA2801  
 C;Superfamily: pol polyprotein  
 Alignment Scores:



Pred. No.:	5,37e-135	Length:	902
Score:	3051.00	Matches:	570
Percent Similarity:	96.45%	Conservatives:	28
Best Local Similarity:	91.94%	Mismatches:	18
Query Match:	66.98%	Indels:	4
DB:	2	Gaps:	2
US-09-610-313B-32 (1-2457) x T01668 (1-902)			
QY	511	ATGATCGGGGATCGCGGGCTTCATCAAGGTGCGCCAGTACGACACGATCTGTATCGAG	570
Db	1	MetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu	20
QY	571	ATCTCGGCAAGAGCCATCGCGCTGATCGCGCCACCCCGTGAACATCATC	630
Db	21	IleCysGlyLysLeuAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle	40
QY	631	GGCCGCAACATGCTACCCAGCTGGGCTGCACCTTGAATCTCCCATCAGCCCAATCGAG	690
Db	41	GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	60
QY	691	ACCTGCGCGTGAAGCCCGCATGAGCGCGCCCAAGGTGAAGAGTGGCCCTG	750
Db	61	ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu	80
QY	751	ACCGAGGAGAAGATCAAGCCCTGACCGCATCTCGAGGAGTGGAGAGGAGGCGAAG	810
Db	81	ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGlnLysGlnLys	100
QY	811	ATCACCAAGATCGGCGCCGAGAACCCCTACAAACACCCCGTTCGCCATCAAGAAGAAG	870
Db	101	IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys	120
QY	871	GACAGCACCAAGTGGCGCAAGCTGTGGACTTTCGCGAGCTGAACAAGCGCACCCAGGAC	930
Db	121	AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp	140
QY	931	TTCTGGAGGTGACGTGGGATCCCCACCCCGCGCGCTGAAGNAGNAGAGCGCTG	990
Db	141	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	160
QY	991	ACCGTGTGGAGTGGGGGACCCCTACTTCAGCTGCCCTCGACGAGGACTTCGCGAAG	1050
Db	161	ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys	180
QY	1051	TACACGCTTCACCATCCCCAGCATCAACAACGAGACCCCGGATCGCTTACCACTAC	1110
Db	181	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr	200
QY	1111	AACTGTCTGCCAGGCTGGAGGGCAGCCCGACGATCTTCAGAGCATGACCAAG	1170
Db	201	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	220
QY	1171	ATCTCTGGAGCCCTTCGCGCCCGCAACCCCGAGATCTGATCTACCAAG-----GCCCC	1224
Db	221	IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp	240
QY	1225	CTGTACGTGGGAGGACCTGAGATTCGCCAGCACCCCGCAAGATCGAGAGTTCGCG	1284
Db	241	LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuArg	260
QY	1285	AGGACCTCTGCTGGTGGGGCTTCAACCCCGGACAGAACGACCAAGAGGAGGCCCC	1344
Db	261	GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	280
QY	1345	TTCTCTG-----CCCATCGAGTGCACCCCGACAGTGCACCGTGCAGCCCATCGAGCTG	1398
Db	281	PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu	300
QY	1399	CCCGAGAAGGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAAGTGG	1458
Db	301	ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	320

QY	1459	GCCAGCCAGATCTACCCCGGCATCAAGGTGGCGCCAGCTGTGCAAGCTGTGCGCGCGCC	1518
Db	321	AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla	340
QY	1519	AAGCCCTGACGACATCGTCCCTGACCCAGGAGGCGAGCTGGAGTGGCCGAGAAC	1578
Db	341	LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn	360
QY	1579	GCGAGATCTTCGCGAGCCCGTCGCGCTGTACTACGACCCCGACCAAGACCTGGTG	1638
Db	361	ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle	380
QY	1639	GCCAGATCCAGAAGCAGGCGCACACAGTGGACCTACCAAGATCTACCAAGAGCCCTTC	1698
Db	381	AlaGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGlnGlnTyr	400
QY	1699	AAGAACCTGAAGACCGCAAGTACGCCAAGATGGCACCGCCACCAACAGACGTGAAG	1758
Db	401	LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys	420
QY	1759	CAGCTGACCGAGCGCGTCAGAAATCCCATGGAGAGCATCGTATCTGGGCGAAGACC	1818
Db	421	GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr	440
QY	1819	CCCAAGTTCCGCTCCCATCCAGAGGAGACCTGGGAGACCTGTGTGACCGACTACTGG	1878
Db	441	ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp	460
QY	1879	CAGCCACCTCGATCCCGAGTGGAGTTCGTGAACACCCCCCTCGTGAAGCTGTGG	1938
Db	461	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp	480
QY	1939	TACCACTGGAGAGGAGGCCCATCATCGCGCGCGAGACTTCTAGCTGGACGCGCGCCG	1998
Db	481	TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla	500
QY	1999	AACCGGAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGCGCGCGGAGAGATC	2058
Db	501	AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal	520
QY	2059	GTGAGCTGACCGGAGACCAACCAAGAGACCGAGTGCAGGCGCATCCAGCTGCGCCCTG	2118
Db	521	ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu	540
QY	2119	CAGGACCGCGGAGGAGGTGAACATCGTACCGACGACGACGATCGCTGGGCGATCATC	2178
Db	541	GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle	560
QY	2179	CAGGCGGACCGCACAGAGCGAGCGAGCTGGTGAACCAAGATCATCGAGCGCTGATC	2238
Db	561	GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGlnLeuIle	580
QY	2239	AAGAGGAGAAGTGTACTCTGAGTGGTGGTCCCGCCCAAGGGCATCGGCGGCAACGAG	2298
Db	581	GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyAsnGlu	600
QY	2299	CAGATCGACAGCTGGTGAAGGCGATCGCAAGGTGCTGTTCTCGACGCGCATCGAT	2358
Db	601	GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp	620
RESULT 12			
B47175			
reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1			
C:Species: human immunodeficiency virus type 1, HIV-1			
C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004			
C:Accession: B47175			
R:Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.			
Proc.Natl. Acad. Sci. U.S.A. 90, 25-29, 1993			
A:Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in th			
A:Reference number: A47175; MUID:93126353; PMID:7678340			
A:Accession: B47175			
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra			
A:Molecule type: DNA			

A;Residues: 1-559 <MOH>  
 A;Cross-references: UNIPROT:Q9PXX1  
 A;Note: sequence extracted from NCBI backbone (NCBIP:122099)  
 C;Superfamily: pol polyprotein

## Alignment Scores:

Pred. No.:	5, 7e-120	Length:	559
Score:	2728.00	Matches:	510
Percent Similarity:	95.35%	Conservative:	23
Best Local Similarity:	91.23%	Mismatches:	22
Query Match:	59.89%	Indels:	4
DB:	2	Gaps:	2

US-09-610-313B-32 (1-2457) x B47175 (1-559)

```

QY 673 CCCATCAGCCCGCCGAGACCGTCCCGTGAAGCTGAAGCCCGGCGATCGAGCGGCCCAAG 732
DB 1 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20
QY 733 GTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAG 792
DB 21 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu 40
QY 793 ATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGAGACCCCTACCAACACCCCGTG 852
DB 41 MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60
QY 853 TTCCGCCATCAAGAGAGGAGCAGACCAAGTGGCGGAGCTGGTGGACTTCCCGGAGCTG 912
DB 61 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 80
QY 913 AACAGCGCCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGCCACCCCGCGGCTG 972
DB 81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100
QY 973 AAGAGAGAGAGCGCTGACCGTCTGACGTGGAGCTGGGCGAGCGCTACTTCAGGTCGCCCTG 1032
DB 101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120
QY 1033 GACGAGACTTCCCAAGTACACCGCTTACCATCCCGACATCAACCAACGAGACCCCG 1092
DB 121 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140
QY 1093 GGCATCCGCTACCAAGTACAAGTCTGCCAGGCTGGAGGCGAGCCCGACGATCTTC 1152
DB 141 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160
QY 1153 CAGAGCATGACCAAGATCTTGAGCCCTTCCGCGCCGCGCAACCCCGAGATCGTGATC 1212
DB 161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIle 180
QY 1213 TACCAG-----GCCCGCTGTAGTGGCGAGCGACCTGGAGATCGCGCCAGCCCGGCC 1266
DB 181 TyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr 200
QY 1267 AAGATCAGGAGCTGCGCAAGACACTGCTGGCTGGGGCTTCCACACCCCGCGCAAGAAAG 1326
DB 201 LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLys 220
QY 1327 CACAGAGAGGAGCCCGCTTCCTG-----CCCATCGAGCTGCACCCCGCAAGTGAAC 1380
DB 221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240
QY 1381 GTGCAGCCCATCGAGCTCCCGAGAGAGAGCTGGACCTGGACCTGACGATCCAGCATCCGAGCTG 1440
DB 241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260
QY 1441 GTGGCGAAGCTGAACCTGGCGCAGCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTC 1500
DB 261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuLys 280
QY 1501 AAGTGTGTCGCGCGCCAGGCGCTGACCGACATCGTCCCTGACCGGAGGCGCGAG 1560

```

```

DB 281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGlu 300
QY 1561 CTGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGCGAGCCCGGTGTACTACGAC 1620
DB 301 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp 320
QY 1621 CCCAGCAAGACCTGGTGGCGGAGATCCAGAAAGCAGGCGCCACGACCATGTGACCTACCAG 1680
DB 321 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGln 340
QY 1681 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGCGAGTACGCAAGATCGGACCCGCC 1740
DB 341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360
QY 1741 CACACCAACACGCTGAAGCAGCTCACGAGGCGCTGCAGAAAGATCGCCATCGAGAGCATC 1800
DB 361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380
QY 1801 GTGATCTGGGCAAGACCCCAAGTTCGGCTGCCATCCAGAGGAGACCTGGGAGACC 1860
DB 381 ValIleTrpGlyLysThrProArgPheLysLeuProIleGlnLysGluThrTrpGluThr 400
QY 1861 TGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCC 1920
DB 401 TrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420
QY 1921 CCCCTGCTGAAGCTGTGTACCGACTGAGAGAGAGCGCCATCATCGGCGCGAGACCTTC 1980
DB 421 ProLeuValLysLeuTrpTyrGlnLeuLysGluProIleValGlyAlaGluThrPhe 440
QY 1981 TACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGTGACCCGAC 2040
DB 441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460
QY 2041 CGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGACCGAGACCGAGCTGAG 2100
DB 461 LysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeuGln 480
QY 2101 GCATCCAGCTGGCGCTGCGAGGACGCGAGCGAGTGAACATCGTGACCGAGCGCCAG 2160
DB 481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500
QY 2161 TACGCTCGGCGCATCATCCAGGCGCCAGCCCGAGAGCGAGAGCGAGCTGGTGAACCGAG 2220
DB 501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520
QY 2221 ATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGTGGTCCCGCCACAG 2280
DB 521 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540
QY 2281 GGATCGGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGCGCATCCGCAAGGTG 2337
DB 541 GlyIleGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleIleArgLysVal 559

```

## RESULT 13

A47175

reverse transcriptase, AZT-resistant variant - human immunodeficiency virus type 1  
 C;Species: human immunodeficiency virus type 1, HIV-1  
 C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C;Accession: A47175  
 R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993  
 A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the  
 A;Reference number: A47175; PMID:93126353; PMID:7678340  
 A;Accession: A47175  
 A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A;Molecule type: nucleic acid  
 A;Residues: 1-559 <MOH>  
 A;Cross-references: UNIPROT:Q9PXX2  
 A;Note: sequence extracted from NCBI backbone (NCBIP:122100)  
 C;Superfamily: pol polyprotein  
 Alignment Scores:

Pred. No.:	4.88e-119	Length:	559
Score:	2708.00	Matches:	506
Percent Similarity:	95.17%	Conservatives:	26
Best Local Similarity:	90.52%	Mismatches:	23
Query Match:	59.45%	Indels:	4
DB:	2	Gaps:	2
US-09-610-313B-32 (1-2457) x A47175 (1-559)			
QY	673	CCCATCAGCCCCATCGAGACCCCTGCGTGAAGCCCGGCGATGAGCGGCCCAAG	732
Db	1	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	20
QY	733	GTGAGCAGTGGCCCTGACCGAGAGATCAAGGCCCTGACCGCATCTGCGAGGAG	792
Db	21	ValLysGlnTrpProLeuThrGluLysIleLysAlaLeuValGluLysCysThrGlu	40
QY	793	ATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCGCGAAGCCCTACAAACCCCGCTG	852
Db	41	MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal	60
QY	853	TTCCGCATCAAGAAGAGACAGCACCAAGTGGCGAAGCTGGTGAGCTTCCGCGAGCTG	912
Db	61	PheAlaIleLysLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu	80
QY	913	AACAGCCGACCCAGGACTTCTGGGAGTGACCTGGGCGATCCCGCCCGCGCGCTG	972
Db	81	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	100
QY	973	AAGAAGAAGAAGCGCTGACCGTGTGGACGTGGCGGCGCGCTACTTACGCGTGGCCCTG	1032
Db	101	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	120
QY	1033	GACGAGGACTTCCGAAATPACACCGCTTACATCCCGCAGCATCAACAAACGAGACCC	1092
Db	121	AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro	140
QY	1093	GCATCCGCTACGATGACAACTGCTGCCAGGGCTGGAAGGGCAGCCCGAGCTCTTC	1152
Db	141	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	160
QY	1153	CAGAGCAGCATCACCAAGATCTCGAGCCCTTCCCGCGCCGCAACCCCGAGATCGTGATC	1212
Db	161	GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspMetValIle	180
QY	1213	THACAG-----GCCCGCTGTACGTGGCAGCGACTGGAGATCGGCCAGACCGCGCC	1266
Db	181	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr	200
QY	1267	AGATCGAGGAGCTGGCAAGCACCTGCTGGCTGGGCTTCACCCACCCCGACAGAGAG	1326
Db	201	LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys	220
QY	1327	CACCAGAAGGAGCCCGCTTCCCTG-----CCCATCGAGCTGCACCCCGCAAGTGGACC	1380
Db	221	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	240
QY	1381	GTGAGCCCATCAGCTGCCGAGAGAGAGAGCTGACCGTGAACGACATCCAGAAAGCTG	1440
Db	241	ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu	260
QY	1441	GTGGCGAGCTGAACCTGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGCTGTGC	1500
Db	261	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys	280
QY	1501	AAAGTGTCTCGGGGGCCCAAGGCCCTGACGACATCGTGCCTGACCGAGGAGGCCGAG	1560
Db	281	LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluGluAlaGlu	300
QY	1561	CTGAGCTGGCGCAGAACCGCGAGATCTGTGGCGGAGCCCGTGCACGCGGTGTACTACGAC	1620
Db	301	LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTrpAsp	320

QY	1621	CCAGACGAGACCTGGTGGCCGAGATCCAGAAGCAGGGGCCACGACCGTGGACCTACCAG	1680
Db	321	ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGln	340
QY	1681	ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTAGTCGCAAGATGCGCACCGC	1740
Db	341	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla	360
QY	1741	CACACCAAGAGCTGAGACAGCTGACCGAGCGCTGCAAGATCGCATCGAGAGCATC	1800
Db	361	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle	380
QY	1801	GTGATCTCGGCAAGACCCCAAGTTCGCTGCCCATCCAGAGAGACCTGGGAGACCC	1860
Db	381	ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla	400
QY	1861	TGGTGACCGCACTACTGGCAGGCCACCTGGATGCCGAGTGGGAGTTCGTGAACACCC	1920
Db	401	TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	420
QY	1921	CCCTGGTGAAGCTGGTACCTGACCTGAGAGGAGCCCATCATCGCGCGCGAGACCTTC	1980
Db	421	ProLeuValLysLeuTrpTyrGlnLeuGlnLysGluProIleValGlyAlaGluThrPhe	440
QY	1981	TACGTGGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC	2040
Db	441	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	460
QY	2041	CGGCGCGCGCAAGATCGTGAGCTGACCGAGACCCCAACCAAGAACCGAGCTGCGAG	2100
Db	461	LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln	480
QY	2101	GCCATCCAGCTGGCCCTCAGACAGCGCGCAGCGAGTGAACATCGTGACCGACAGCCAG	2160
Db	481	AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln	500
QY	2161	TACGCGCTGGGCATCATCCAGCGCCAGCCGACAGAGCAGAGCGAGCTGGTGAACACAG	2220
Db	501	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln	520
QY	2221	ATCATCGAGCAGCTGATCAAGAAGAGAGTGTACTCTGAGCTGGTGGTCCCGCCCAAG	2280
Db	521	IleIleGluGluLeuIleLysLysGlyLysValTyrLeuAlaTrpValProAlaHisLys	540
QY	2281	GGCATCGCGCGCAAGCAGCATCGACAACTGGTGGAGCGGCGCATCCGCAAGCTG	2337
Db	541	GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal	559
RESULT 14			
S46347			
pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)			
C:Species: simian immunodeficiency virus SIVagm			
A:Variety: isolate SAB-1			
C:Date: 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999			
C:Accession: S46347			
R:Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.;			
EMBO J. 13, 2935-2947, 1994			
A:Title: Mosaic genome structure of simian immunodeficiency virus from West African green			
A:Reference number: S46335; MUID:94298785; PMID:8026477			
A:Accession: S46347			
A:Status: nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-1039 <JIN>			
A:Cross-references: EMBL:U04005; NID:9466229; PIDN:AAA21505.1; PID:9466231			
A:Experimental source: isolate SAB-1; sabaeus monkey			
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993			
C:Genetics: this reading frame extends between two stop codons and does not begin with a sta			
A:Gene: pol			
C:Superfamily: pol polyprotein			
Alignment Scores:	2.61e-100	Length:	1039
Pred. No.:			

[illegible]

```
Db 678 SerLysValAsnIleIleThrAspSerGlnTyrAlaMetGlyIleAlaGlyLeuPro 697
QY 2191 GACAGAGCGGAGCGAGCTGTGTAACACGATCATCGACGACTCATCAAGAGGAGAG 2250
Db 698 ThrGluSerAspAsnAsnIleValGlnGlnIleIleGluLeuIleGlyLysGluAla 717
QY 2251 GTGTACCTGAGCTGGTGGCCGCCACCAAGGCGCATCGCGGCAACGAGCATCGACAAG 2310
Db 718 ValTyrIleAlaTrpValProIleHisLysGlyValGlyGlnGluIleAspLys 737
QY 2311 CTGGTGAGCAAGGCGCATCGCAAGGTGCTGTTCTCGGCGGCATCGAT 2358
Db 738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753

RESULT 15
S53092
pol polyprotein - human immunodeficiency virus type 2
C:Species: human immunodeficiency virus type 2, HIV-2
C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C:Accession: S53092
R:Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.
submitted to the EMBL Data Library, March 1995
A:Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu
A:Reference number: S53091
A:Accession: S53092
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1055 <BEC>
A:Cross-references: UNIPROT:O73194; EMBL:Z48731
C:Superfamily: pol polyprotein
C:Keywords: polyprotein

Alignment Scores:
Pred. No.: 6,37e-94 Length: 1055
Score: 2170.00 Matches: 413
Percent Similarity: 70.21% Conservative: 129
Best Local Similarity: 53.50% Mismatches: 187
Query Match: 47.64% Indels: 43
Db: 2 Gaps: 11

US-09-610-313B-32 (1-2457) x S53092 (1-1055)
QY 156 GGGCTGCTGGAAGTCGGCGAAGGAGGCCACCAGATGAAGGACTGCAC---CGAGCGCA 212
Db 3 GlyLeuLeuGluMetTrpGln-----AspArgThrTyrHisGlyLysValPro 18
QY 213 GCCCAACTT-----CTTCG 227
Db 19 ArgGlnThrGlyGlyPhePheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38
QY 228 CGAGGACTGGCTTCCCCCGGCAAGGCCCGCGAGTTCCCGAGGACAGAA-----281
Db 39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerGly 58
QY 282 CGCGCCCAACAGCCCGGCGCA-----GTCGAGGTGGCGGCGCAACCCCG 335
Db 59 ProThrGlyGluIleTyrAlaAlaArgGluLysAlaGluGlyAlaGluArgGluThrIle 78
QY 336 CAGCGAGCGCGCGC-----CGAGCGCA-----GGCGACCCCTCAACTT-----374
Db 79 GlnArgGlyAspArgGlyLeuAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98
QY 375 -----CCCCAGATCACCTGTGGCGAGCCCGCCCTGTGTGAGCATCAA 416
Db 99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValValThrAlaHi 118
QY 417 GTGGCGGCGCAGATCAAGGAGCCCTGCTGACACCGCGCGCGACACCGCTCGGA 476
Db 118 sIleGluGlnProValGluValLeuLeuAspThrGlyAlaAspSerIleValAl 138
QY 477 GGAGATGAGCTCGCGCGCAAGTGGAAAGCCCAAGATGATCGCGCGGCTTCAT 536
Db 138 aglyIleGluLeuGlySerAsnTyrSerProLysIleValGlyIleGlyPheIle 158
```

```
QY 537 CAAGGTGCGCCAGTAGCACCATCTCTGATCGAGATCTGCGCGCAAGAGCCCATCGGCAC 596
Db 158 eAsnThrLysGluTyrLysAsnValGluIleGluValLeuGlyLysValArgAlaTh 178
QY 597 CGTGTGATCGCGCCCGCCACCCCGTGAACATCATCGCGCGCAACATCTGTGACCCAGCTGG 656
Db 178 rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuGl 198
QY 657 CTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGCGTGAAGTGAAGCCCGG 716
Db 198 yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProGl 218
QY 717 CATGGACGCGCCCAAGGTGAAGCGAGTGGCCCTGCACCGAGGAGAGATCAAGGCCCTGAC 776
Db 218 yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeuLys 238
QY 777 CGCCATCTGCGAGGAGATGAGAGAGGAGGCAAGATCAACCAAGATCGCGCCCGGAGAACCC 836
Db 238 sGluIleCysGluLysMetGluArgGluGlyGlnLeuGluAlaProProThrAsnPr 258
QY 837 CTACAACACCCCGTGTTCGCATCAAGAAGAAGACAGCACCAAGTGGCGGCAAGCTGT 896
Db 258 ofYrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysTrpArgMetLeuI 278
QY 897 GGACTTCCGCGAGCTGAACAAAGCGCACCGAGACTTCTGGGAGGTGCGAGTGGGCATCCC 956
Db 278 eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyIlePr 298
QY 957 CCACCCCGCGCTGGAAGAAGAAGAGCGTGACCGTGTGGACGTGGGCGAGCGCTTA 1016
Db 298 oHisProAlaGlyLeuAlaLysLysArgArgIleThrValLeuAspValGlyAspAlaTy 318
QY 1017 CTTGAGCTGCGCTGAGAGAGACTTCCGCAAGTACACCGCTTTCATCCATCCCGAGCAT 1076
Db 318 rPheSerIleProLeuHisGluAspPheArgGlnTyrThrAlaPheThrLeuProSerVa 338
QY 1077 CAACAACGAGACCCCGCATCCGCTACAGTACACAGTGTGCTGCCAGGCGGTGAAGGG 1136
Db 338 lAsnAsnAlaGluProGlyLysArgTyrIleTyrLysValLeuProGlnGlyTrpLysGl 358
QY 1137 CAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCGGAGCGCTTCCGCGCCCGCAA 1196
Db 358 ySerProAlaIlePheGlnTyrThrMetArgGlnValLeuGluProPheArgLysAlaAs 378
QY 1197 CCGCGAGATCGTATC-----TACCAGCGCCCTGTACGTGGCGACGACCTCGAGAT 1250
Db 378 nGlnAspValIleIleIleGlnTyrMetAspAspIleLeuIleAlaSerAspArgThrAs 398
QY 1251 CGGCCAGCACCGGCCAAGATCGAGGAGCTCGCAAGCACCTGTGCTGCGTGGGCGTTCAC 1310
Db 398 pLeuGluHisAspArgValValLeuGlnLeuLysGluLeuLeuAsnSerLeuGlyPheSe 418
QY 1311 CACCCCGCACAGAACACCAAGAGAGCGCCCTTC-----CTGCCCATCGAGCTGCA 1364
Db 418 rThrProAspGluLysPheGlnLysAspProProTyrArgTrpMetGlyTyrGluLeuTr 438
QY 1365 CCGCGACAAAGTGGACCGCTGTCAGCCCATCGAGTGCCTGCGAGAGAGAGAGTGGACCGTGA 1424
Db 438 pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysGluValTrpThrValAs 458
QY 1425 CGACATCCAGAGCTGTGGCGCAAGCTGAATGGCGCCAGCCAGATCTACCCCGCATCAA 1484
Db 458 nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyrProGlyIleLys 478
QY 1485 GTGCGCGCAGCTGTCAAGCTGCTCGCGCGCGGCCAAGGCCCTGACCGCATCGTCCCT 1544
Db 478 sThrLysHisLeuCysArgLeuIleArgGlyLysMetThrLeuThrGluGluIleGlnTr 498
QY 1545 GACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGGAGCCCGTGA 1604
Db 498 pThrGluLeuAlaGluAlaGluLeuGluGluAsnArgValIleLeuSerGlnGluGlnGl 518
```





***This Page Blank (uspio)***



DR	InterPro:	IPR000721;	Gag_p24.
DR	InterPro:	IPR001037;	Integrase_C.
DR	InterPro:	IPR003308;	Integrase_Zn_N.
DR	InterPro:	IPR001995;	Peptidase_A2
DR	InterPro:	IPR009007;	Pept_Aspartic.
DR	InterPro:	IPR001969;	Pept_Asp_AS.
DR	InterPro:	IPR010999;	Retrovir_matrix.
DR	InterPro:	IPR000071;	Retrovir_p17.
DR	InterPro:	IPR008916;	Retrov_capsid_C.
DR	InterPro:	IPR008919;	Retrov_capsid_N.
DR	InterPro:	IPR002156;	RNaseH.
DR	InterPro:	IPR001584;	Rve.
DR	InterPro:	IPR000477;	RVTse.
DR	InterPro:	IPR010659;	RVT_connect.
DR	InterPro:	IPR010661;	RVT_thumb.
DR	InterPro:	IPR001878;	Znf_CCHC.
DR	Pfam:	PF00540;	Gag_p17; 1.
DR	Pfam:	PF00607;	Gag_P24; 1.
DR	Pfam:	PF02022;	Integrase_Zn; 1.
DR	Pfam:	PF00075;	RNaseH; 1.
DR	Pfam:	PF00665;	rve; 1.
DR	Pfam:	PF00077;	RVP; 1.
DR	Pfam:	PF00078;	RVT_1; 1.
DR	Pfam:	PF06815;	RVT_connect; 1.
DR	Pfam:	PF06817;	RVT_thumb; 1.
DR	Pfam:	PF00098;	zf-CCHC; 2.
DR	SMART:	SM00343;	ZnF_C2HC; 2.
DR	PROSITE:	PS00141;	ASP_PROTEASE; 1.
DR	PROSITE:	PS00175;	ASP_PROT_RETROV; 1.
DR	PROSITE:	PS0158;	ZF_CCHC; 2.
DR	KW	AIDS;	Aspartyl protease; Core protein; Hydrolyase; Polyprotein;
DR	KW	Protease;	RNA-directed DNA polymerase; Transferase.
SQ	SEQUENCE	1427 AA;	161059 MW; F3B0C741A1EFB10A CRC64;

Alignment Scores:  
 Pred. No.: 3.67e-155 Length: 1427  
 Score: 4097.00 Matches: 777  
 Percent Similarity: 98.98% Conservative: 1  
 Best Local Similarity: 98.85% Mismatches: 4  
 Query Match: 89.95% Indels: 4  
 DB: 2 Gaps: 2

US-09-610-313B-32 (1-2457) x Q9WF92 (1-1427)

Qy	13	ATGGCGGAGGCCTGAGCCAGGCCACCAAGCCAACTCTTGATCAGCGCAGCAACTTC	72
Dd	360	LeuAlaGluAlaMeSerGlnAlaThrSerAlaAenIleLeuWeGInArgSerAenPhe	379
Qy	73	AAGGGCCCCAAGCGCATCATCAAGTGCTTTAACTGCGGCAAGGAGGCCACATCGCCGC	132
Dd	380	LysGlyProLysArgGlleilellysCysPheAenCysGlyLysGluGlyHislealaArg	399
Qy	133	AACTGCGCGCCCCCGCAAGAAGGGCTGTGTGAAGTGGCGGAAGAGGGCCACCAGATG	192
Dd	400	AenCysArgAlaProArgLysLysGlyCystrpLysCysSgLYsGluGlyHisGlnMet	419
Qy	193	AAGACTGTCRCGAGCGCCAGCCCAACTCTTCGGGAGGACCTGGCTTCCCCCAGGGC	252
Dd	420	LysaspCysftrGiuArgGlnAlaAenPhePheArgGluAspLeuAlaPheProGlnGly	439
Qy	253	AAGGCCCGCGATTCCCCAGCGAGCAGAACCGCGCAAACAGGCCCAACAGCCGCGAGCTG	312
Dd	440	LysAlaArgGluPheProSerGluGlnAenArgAlaAenSerProThrSerArgGluLeu	459
Qy	313	CAGGTGCGCGCGCAACAACCCCGCAGCGAGGCCGCGCGAGCCGACCCCTGTAAC	372
Dd	460	GlnValArgGlyAspAenProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAen	479
Qy	373	TTCGCCGAGATCACCTGTGGCAGCGCCCCTGTGTGAGCATCAAGGTGGCGCGCCAGTC	432
Dd	480	PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysVaIGlyGlyGlnIle	499
Qy	433	AAGAGGGCCCTGCTGCACACCGCGCCCGCAACGACACCGTCTGGAGAGATGAGCTGCC	492



Db 360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379  
:::  
Qy 73 AAGGCGCCAGCGCATCATCAAGTGCCTCAACTCGGCGCAAGAGGGCCACATCGCCCGC 132  
Db 380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
Qy 133 AACTCGCGCGCCCGCAAGAGAGGCTGCTGAAGTGGCGCAAGAGGGCCACAGATG 192  
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
Qy 193 AAGACTGCACCGAGCGCCAGCCAACTTCTTCGCGAGGACCTGCGCTTCCGCCAGGGC 252  
Db 420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439  
Qy 253 AAGGCGCGGAGTTCCCGAGCGAGAGAACCGCGCCCAACAGCCCCACAGCGCGAGCTG 312  
Db 440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459  
Qy 313 CAGGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCGAGCGCCAGCGCACCTGAAC 372  
Db 460 GlnValArgLysAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479  
Qy 373 TTCGCCACGATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGCGCAGATC 432  
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499  
Qy 433 AAGGAGCGCTGCTGGACACCGGCGCGACGACACCGTGTGGAGGAGATGAGCCGTGCC 492  
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519  
Qy 493 GSCAAGTGAAGCCCAAGATGATCGCGCATCGCGGCTTCATCAAGTGGCGCGAGTAC 552  
Db 520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539  
Qy 553 GACCAGATCCTGATCGAGATCTGCGCAAGAGGCGCATCGGCACCGTGTGATCGGCCCC 612  
Db 540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559  
Qy 613 ACCCGGTGAACATCATGCGCGCGCAACATGCTGACCGAGCTGGGTGACCCCTGAACCTC 672  
Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579  
Qy 673 CCCATCAGCCCATCGAGACCTGCGCGTGAAGCTGAAGCCGGCATGAGCGGCCCAAG 732  
Db 580 ProfileSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
Qy 733 GTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619  
Qy 793 ATGGAGAGGAGGGCAAGATCACCAAGATCGCGCGCGAGAACCCCTACACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
Qy 853 TTCGCCATCAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTG 912  
Db 640 PheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAAAGCGCACCGAGACTTCTGGGAGGTGCAAGTGGGATCCCCACCCCGCGCGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACTTACGCGTGGCCCTG 1032  
Db 680 LysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGGACTTCGCAAGTACACCCCTTCCATCCCGCAGCATCAACAAACGAGACCC 1092  
Db 700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
Qy 1093 GGCATCCCGTACAGTACAACTGTGTCGCCAGGGCTGGAAGGGCAGGCCCGCAGCATCTTC 1152  
:::  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
Qy 1153 CAGAGCAGCATGACCAAGATCCTGAGGACCTTCGCGCGCCCAACCCCGAGATCGTGTATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759  
Qy 1213 TACCAG-----GCCCGCTGTACGTGGCGAGCACTGTGGAGATCGGCACGCGCCG 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1267 AAGATCAGGAGCTGCGCAAGCAGCATCTGCTGCGTGGGGCTTCACCAACCCCGCAGAGAG 1326  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
Qy 1327 CACCAGAGGAGCGCCCTTCTCTG-----CCCATCGAGCTGCACCCCGACAAGTGGAC 1380  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1381 GTGCAGCCCATCGAGCTGCGCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440  
Db 820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1441 GTGGGCAAGCTGAACCTGGGCGCAGCATCTACCCCGGCATCAAGTGTGGCGCAGCTGTC 1500  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
Qy 1501 AAGCTGTGCGCGCGCGCAAGCGCTGACCGACATCGTCCCTGACCGAGGAGGCCGAG 1560  
Db 860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
Qy 1561 CTGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACCGCGTGTACTACGAC 1620  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899  
Qy 1621 CCCAGAGAGACCTGCTGCGCGAGATCCAGAGCGAGGCCACGACCTGACCTACACAG 1680  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTAGTACGCCAAGATGGCAGCCG 1740  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCGCGTGCAGAGATCGCCATGGAGAGCATC 1800  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
Qy 1801 GTGATCTGGGCGAGACCCCAAGTTCGCTGCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCG 1920  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1921 CCCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTC 1980  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
Qy 1981 TACGTGAGCGCGCGCCCAACCGCGAGACCGAGATCGGCAAGGCCGCGGTCTGTGACCCG 2040  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2041 CGGGCGCGCGAGAGATCGTGGCTGACCGAGACCCAGACCCACCAAGACCGAGCTGGAG 2100  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
Qy 2101 GCCATCCAGCTGGCCCTGCGAGACAGCGCGCAGCGAGGTGAACATCGTGACCGAGCCAG 2160  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy 2161 TACGCCCTGGGCGATCATCCAGGCCCGAGCCCGACAGAGCGAGCGAGCTGTGTACACG 2220  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099

Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
Qy 1153 CAGAGCAGCATGACCAAGATCCTGAGGACCTTCGCGCGCCCAACCCCGAGATCGTGTATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759  
Qy 1213 TACCAG-----GCCCGCTGTACGTGGCGAGCACTGTGGAGATCGGCACGCGCCG 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1267 AAGATCAGGAGCTGCGCAAGCAGCATCTGCTGCGTGGGGCTTCACCAACCCCGCAGAGAG 1326  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
Qy 1327 CACCAGAGGAGCGCCCTTCTCTG-----CCCATCGAGCTGCACCCCGACAAGTGGAC 1380  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1381 GTGCAGCCCATCGAGCTGCGCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440  
Db 820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1441 GTGGGCAAGCTGAACCTGGGCGCAGCATCTACCCCGGCATCAAGTGTGGCGCAGCTGTC 1500  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
Qy 1501 AAGCTGTGCGCGCGCGCAAGCGCTGACCGACATCGTCCCTGACCGAGGAGGCCGAG 1560  
Db 860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
Qy 1561 CTGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACCGCGTGTACTACGAC 1620  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899  
Qy 1621 CCCAGAGAGACCTGCTGCGCGAGATCCAGAGCGAGGCCACGACCTGACCTACACAG 1680  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTAGTACGCCAAGATGGCAGCCG 1740  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCGCGTGCAGAGATCGCCATGGAGAGCATC 1800  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
Qy 1801 GTGATCTGGGCGAGACCCCAAGTTCGCTGCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCG 1920  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1921 CCCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTC 1980  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
Qy 1981 TACGTGAGCGCGCGCCCAACCGCGAGACCGAGATCGGCAAGGCCGCGGTCTGTGACCCG 2040  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2041 CGGGCGCGCGAGAGATCGTGGCTGACCGAGACCCAGACCCACCAAGACCGAGCTGGAG 2100  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
Qy 2101 GCCATCCAGCTGGCCCTGCGAGACAGCGCGCAGCGAGGTGAACATCGTGACCGAGCCAG 2160  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy 2161 TACGCCCTGGGCGATCATCCAGGCCCGAGCCCGACAGAGCGAGCGAGCTGTGTACACG 2220  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099





Db 580 ProfileSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
QY 733 GTGAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGCCAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619  
QY 793 ATGGAGAGAGGAGGGAAGATCACCAAGATCGGCCCGCGAAGCCCTCAACAACCCCGCTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
QY 853 TTGCGCATCAAGAAGACAGACCAAGTGGCGCAAGCTGGTGGACTTCGCCAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
QY 913 AACAGCCGACCCAGGACTTCTGGAGGTGACCTGGGCGATCCGCCACCCCGCGGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
QY 973 AAGAAGAAGAAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACTTTCAGCGTGGCCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
QY 1033 GACGAGGACTTCGCAAGTACACCGCTTACCATCCCGAGCATCAACAACGAGACCCCC 1092  
Db 700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
QY 1093 GCATCCCGCTACAGTACAACTGCTGCCGAGGCTGGAAGGGCAGGCCCGAGCATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
QY 1153 CAGAGCAGCATCACCAAGATCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValle 759  
QY 1213 TACCAG-----GCCCGCCCTGACTGGCGACGACCTGGAGATCGGCCAGCACCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
QY 1267 AGATCGAGGAGCTGGGAGACCTGCTGGCTGGGCTTACCAACCCCGAGCAAGAAG 1326  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrValAsnAspIleGlnLys 799  
QY 1327 CACCAAGAGGAGCCCGCTTCCTG-----CCCATCGAGCTGCACCCCGACAAGTGGAGCC 1380  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
QY 1381 GTGAGCCCATCGAGTGGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1440  
Db 820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
QY 1441 GTGGCAGCTGAATGGGCCGACGAGATCTACCCCGGATCAAGGTGGCGGAGCTGTGC 1500  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
QY 1501 AAGCTGCTCGCGGGCCAAAGCCCTGACCGATCGTCCCTGACCGAGGAGGCGCAG 1560  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1561 CTGGAGCTGGCCGAGAACCGCAGATCTCTGCGCGAGCCCGTGCACCGCGGTGTACTACGAC 1620  
Db 880 LeuGluLeuAlaGluAspArgGluIleLeuArgGluProValHisGlyValTyrTrpAsp 899  
QY 1621 CCCAGGAGGACTGGTGGCCGAGATCCAGAGCGAGGCGCACGACGAGTGGACCTACGAG 1680  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
QY 1681 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCGCACCGCC 1740  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
QY 1741 CACACCAACGAGTGAAGCAGCTGACCGAGGCGCGTGCAGAAGATCGCCATGGAGAGCATC 1800  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

QY 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 960 VallietrpglyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
QY 1861 TGGTGGACCGGACTACTGGCAGGCCACCTGGATCCCCCGAGTGGGAGTTCTGTGAACACCCCC 1920  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTyrGluPheValAsnThrPro 999  
QY 1921 CCCCTGTGAAGCTGTGTATCAGCTGGAGAAGGCCCATCATCGGCCCGCAGACCTTC 1980  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
QY 1981 TACGTGGACGCGCGCCCAACCGGAGACCAAGATCGCAAGCGCGGTACGTGACCCGAC 2040  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
QY 2041 CGGGCCCGCGAGAAGATCGTGAGCTGACCGAGACCCACCAACCAAGAGACCGAGCTGCAG 2100  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
QY 2101 GCCATCAGCTGGCCCTGCAGACAGCGGCGAGCGAGTGAACATCGTACCGACAGCCAG 2160  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
QY 2161 TACGCCCTGGGCATCATCCAGCCCGACGACCAAGAGCGAGCGAGCTGGTGAACCCAG 2220  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
QY 2221 ATCATCGAGCAGCTGATCAAGAAGAGAGTGTACTCTGAGCTGGTGGTCCCGCCACCAAG 2280  
Db 1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2281 GGCATCGCGCGGCAAGCAGACATCGAAGCTGTGAGCAAGGCGATCGGCAGGTGCTG 2340  
Db 1120 GlyIleGlyGlyAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139  
QY 2341 TTCCTGGACGGCATCGAT 2358  
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 4  
Q9WF62  
ID Q9WF62 PRELIMINARY; PRT; 1427 AA.  
AC Q9WF62;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_taxid=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndong'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RT Botswana.";  
RL J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110961; AAD17021.1; -.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.



QY 1381 GTGAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTG 1440  
DB 820 ValGlnProIleAsnLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
QY 1441 GTGGGCAAGCTGAACCTGGCCGAGCAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGC 1500  
DB 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyProGlyIleLysValArgGlnLeuCys 859  
QY 1501 AAGCTGTGCGGGGCGCCAGGCCCTGACCGATCGTGGCCCTGACCGAGAGGCCGAG 1560  
DB 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1561 CTGGAGCTGGCGAGAACCGCAGATCTGCGCGAGCCGCTGCACGGGTGTACTACGAC 1620  
DB 880 LeuGluLeuAlaGluAsnArgGluLeuLysGluProValHisGlyValTyTrpAsp 899  
QY 1621 CCCAGCAGGACCTGTGGTGGCAGATCCAGAGCAGGGCCAGCAGCTGGACCTACGAG 1680  
DB 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyGln 919  
QY 1681 ATCTACGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATGGCGACCGCC 1740  
DB 920 IleTyGlnGluProPheLysAsnLeuLysThrGlyLysTyTrpAlaLysMetArgThrAla 939  
QY 1741 CACACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGGACATC 1800  
DB 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle 959  
QY 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTTCCGCTCCATCCAGAGGAGACCTGGGAGACC 1860  
DB 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
QY 1861 TGGTGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCC 1920  
DB 980 TrpTrpThrAspTyTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
QY 1921 CCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTC 1980  
DB 1000 ProLeuValLysLeuTrpTyGlnLeuGluGluProIleValGlyAlaGluThrPhe 1019  
QY 1981 TACGTGGAGCGGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTGCTGACCGCAC 2040  
DB 1020 TyrValAspGlyAlaAsnArgGluThrLysIleGlyLysAlaGlyTyValThrAsp 1039  
QY 2041 CGGGCGCGCAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2100  
DB 1040 ArgGlyArgGluArgIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
QY 2101 GCATCCAGCTGGCCCTGCAGACACCGCGCAGGAGTCAACATCGTACCGACGAGCCAG 2160  
DB 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln 1079  
QY 2161 TACGCGCTGGGCATCATCCAGCGCCCGCAGCAAGAGCGAGCGAGCTGGTGAACCCAG 2220  
DB 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
QY 2221 ATCATCGACGACTGATCAAGAGGAGAGGTGTACTTGAAGTGGTGGTGGCCGCCACAG 2280  
DB 1100 IleIleGluGlnLeuIleLysLysGluArgValTyLeuSerTrpValProAlaHisLys 1119  
QY 2281 GGCATCGGCGCAAGCAGGATCGACAGCTGGTGGCAGGCGATCGCAGAGGTGCTG 2340  
DB 1120 GlyIleGlyGlyLysGlnGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
QY 2341 TTCTGGACGCGATCGAT 2358  
DB 1140 PheLeuAspGlyIleAsp 1145  
RESULT 5  
Q6X4P8  
ID Q6X4P8 PRELIMINARY; PRT; 1429 AA.  
AC Q6X4P8;

DT 05-JUL-2004 (TReMBLrel. 27, Created)  
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)  
DE Gag-pol fusion polypeptide (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviridae; Retroviridae; Lenti-virus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14709249; DOI=10.1089/0892220377181220;  
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birk D.L.,  
RA McCutchan F.E.;  
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
in full genome analysis, is not sustained in subgenomic regions.";  
RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Harris M.E., Birk D.L.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kim B., McCutchan F.E.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Maayan S., Zeira M.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR HSGP; P03367; 1A8G.  
DR GO: GO:0019012; C:virion; IEA.  
DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0008907; F:peptidase activity; IEA.  
DR GO: GO:0008233; F:peptidase activity; IEA.  
DR GO: GO:0004523; F:ribonuclease H activity; IEA.  
DR GO: GO:0003723; F:RNA binding; IEA.  
DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR GO: GO:0016740; F:transferase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0015074; P:DNA integration; IEA.  
DR GO: GO:0006310; P:DNA recombination; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO: GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR00721; Gag\_p24.  
DR InterPro; IPR001037; Integrase C.  
DR InterPro; IPR003308; Integrase Zn N.  
DR InterPro; IPR001995; Peptidase A2.  
DR InterPro; IPR009007; Pept\_Aspartic.  
DR InterPro; IPR001969; Pept\_Asp\_AS.  
DR InterPro; IPR010999; Retrovir matrix.  
DR InterPro; IPR000071; Retrovir p17.  
DR InterPro; IPR008916; Retrov capsid C.  
DR InterPro; IPR008919; Retrov capsid\_N.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR InterPro; IPR010659; RVT connect.  
DR InterPro; IPR010661; RVT thumb.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF02022; Integrase Zn; 1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR Pfam; PF06815; RVT\_connect; 1.  
DR Pfam; PF06817; RVT\_thumb; 1.  
DR Pfam; PF00098; zF-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.



|||||  
978 ThrTrpGluThrTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPhe 997  
|||||  
1909 GTGACACACCCCTCGTGTGAAGCTGGTACAGCTGGAGGAGAGGCCATCATCGGC 1968  
|||||  
998 ValAsnThrProProLeuValLysLeuTyrTrpGlnLeuGluLysGluProIleAlaGly 1017  
|||||  
1969 GCGGAGACCTCTACGTGACCGCGCGCCGACCAACCGGACCAAGATCGGCAAGCCGCG 2028  
|||||  
1018 ValGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1037  
|||||  
2029 TACGTGACCGACCGCGCGCGCAGAGAGTCTGAGCTGACCGACCAACCAACGAGAAG 2088  
|||||  
1038 TyrValThrAspArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLys 1057  
|||||  
2089 ACCGAGCTGCAGCCATCAGCTGCGCTGCGAGGACAGCGGAGCGAGTGAACATCGTG 2148  
|||||  
1058 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1077  
|||||  
2149 ACCGACGCGCAGTACGCTGCGCGCATCATCCAGCGCCGACCGGACGAGCGAGCGAG 2208  
|||||  
1078 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 1097  
|||||  
2209 CTGGTGAACACGATCATCGAGCAGCTGATCAAGAAGAGGAGAGCTGACTGAGCTGGGTG 2268  
|||||  
1098 LeuValAsnGlnIleIleGluGlnLeuIleArgLysGluArgValTyrLeuSerTrpVal 1117  
|||||  
2269 CCGCGCCCAAGGGATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAAGCGGCATC 2328  
|||||  
1118 ProAlaHisLysGlyIleGlyGlnGluGlnValAspLysLeuValSerSerGlyIle 1137  
|||||  
2329 CCGAAGGTGCTCTCTGACCGCATCGAT 2358  
|||||  
1138 ArgLysValLeuPheLeuAspGlyIleAsp 1147  
|||||

## RESULT 6

Q6X4Q6 PRELIMINARY; PRT; 1425 AA.  
AC Q6X4Q6;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Gag-pol fusion polypeptide (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14709249; DOI=10.1089/088922203771881220;  
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birk D.L.,  
RA McCutchan F.E.;  
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
RT in full genome analysis, is not sustained in subgenomic regions.";  
RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Harris M.E., Birk D.L.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kim B., McCutchan F.E.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Maayan S., Zeira M.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -I- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL: AY255824; AAP76529.1; --  
DR HSP; Q72498; 1K6Y.  
DR GO: GO:0019012; C:virion; IEA.  
DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0008907; F:integrase activity; IEA.

DR GO: GO:0008233; F:peptidase activity; IEA.  
DR GO: GO:0004523; F:ribonuclease H activity; IEA.  
DR GO: GO:0003723; F:RNA binding; IEA.  
DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR GO: GO:0016740; F:transferase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0015074; F:DNA integration; IEA.  
DR GO: GO:0006310; F:DNA recombination; IEA.  
DR GO: GO:0006278; F:proteolysis and peptidolysis; IEA.  
DR GO: GO:0016032; F:viral life cycle; IEA.  
DR InterPro: IPR000721; Gag\_P24.  
DR InterPro: IPR001037; Integrase\_C.  
DR InterPro: IPR003308; Integrase\_Zn\_N.  
DR InterPro: IPR001995; Peptidase\_A2.  
DR InterPro: IPR009007; Pept Aspartic.  
DR InterPro: IPR001969; Pept Asp AS.  
DR InterPro: IPR010999; Retrovir\_matrix.  
DR InterPro: IPR000071; Retrovir\_p17.  
DR InterPro: IPR008916; Retrov\_capsid\_C.  
DR InterPro: IPR008919; Retrov\_capsid\_N.  
DR InterPro: IPR002156; RNaseH.  
DR InterPro: IPR001584; Rve.  
DR InterPro: IPR000477; RVTse.  
DR InterPro: IPR010659; RVT\_connect.  
DR InterPro: IPR010661; RVT\_thumb.  
DR InterPro: IPR001878; Znf\_CCHC.  
DR Pfam: PF00540; Gag\_p17; 1.  
DR Pfam: PF02022; Integrase\_Zn; 1.  
DR Pfam: PF00075; RNaseH; 1.  
DR Pfam: PF00665; rve; 1.  
DR Pfam: PF00077; RVP; 1.  
DR Pfam: PF00078; RVT\_1; 1.  
DR Pfam: PF06815; RVT\_connect; 1.  
DR Pfam: PF06817; RVT\_thumb; 1.  
DR Pfam: PF00098; zf-CCHC; 2.  
DR PRINTS: PR00234; HIVMATRIX.  
DR SMART: SM00343; Znf\_C2HC; 2.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
DR PROSITE: PS00175; ASP\_PROT\_RETROV; 1.  
DR PROSITE: PS00158; ZF\_CCHC; 2.  
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
KW Zinc-finger.  
FT NON TER 1  
SQ SEQUENCE 1425 AA; 161115 MW; 782315D4D5305E18 CRC64;

Alignment Scores:  
Pred. No.: 8.2e-148 Length: 1425  
Score: 3912.00 Matches: 730  
Percent Similarity: 96.69% Conservative: 30  
Best Local Similarity: 92.88% Mismatches: 22  
Query Match: 85.88% Indels: 4  
Gaps: 2  
DB: 2

US-09-610-313B-32 (1-2457) x Q6X4Q6 (1-1425)

QY 13 ATGGCGGAGGCGCATGAGCGGCGCACGAGCGCCACATCTGATGCGAGCGCAACTTC 72  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 358 LeuAlaGluAlaMetThrGlnAlaAsnAsnThrHisIleMetMetGlnArgAsnAsnPhe 377  
QY 73 AAGGCGCCCAAGCGCATCATCAAGTGTCTTAACCTCGGCAAGAGGCGCCACATCGCCGCG 132  
|||||  
Db 378 LysGlyProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 397  
QY 133 AACTCG 192  
|||||  
Db 398 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 417  
QY 193 AAGGACTGCACCGAGCG 252







640 PheAlaIleLysLysLysAspSerThrIlystrpàrgLysLeuValAspPheAraGluLeu 659  
913 AACAGGCGACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCGACCCCGCGGCTG 972  
660 AsnLysargThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
973 AAGAAGAGAGAGCGTGCCTGCTGAGCTGGCGGACGCTACTTTCAGGCTGCCCTG 1032  
680 LysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
1033 GACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGACCATCAACAACGAGACCC 1092  
700 AspGluGlyPheAraGlyLysThrAlaPheThrIleProSerIleAsnAlaThrPro 719  
1093 GGATCCGCTACCACTACAACTGCTGCCAGGCTGGAGGCGAGCCCGACGATCTTC 1152  
720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
1153 CAGAGCATGATGACCAAGATCTCGAGCCCTTCCGCGCCGACCAACCCCGAGATCGTATC 1212  
740 GlnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759  
1213 TACCAG-----GCCCGCTTCTGCTGGCGAGCGACCTGGAGATCGGCGACCGCGCC 1266  
760 TyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
1267 AAGATCGAGGAGTGGCGAAGCATCTGCTGGCTGGGGCTTCCACACCCCGCAAGAAG 1326  
780 LysIleGluLeuLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799  
1327 CACCAGAGGAGCCCGCTTCTCTG-----CCCATCGAGTGCACCCCGCAAGTGAC 1380  
800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
1381 GTGCAGCCCATCGAGTGGCGAAGAGAGTGGACCTGACCGTGAACGATCCAGAGCTG 1440  
820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
1441 GTGGCAAGCTGAACTGGCGGAGCGAGCATCTACCCCGCATCAAGTGGCGACCTGTC 1500  
840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuLys 859  
1501 AAGTGTCTGGCGGCGCAAGGCGCTGACCGACATCGTGCCTTCCCGTGGAGGCGCGAG 1560  
860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879  
1561 CTGAGCTGGCGGAGACCGGAGATCTCGCGAGCGCGCTGACCGCTGTACTAGAC 1620  
880 LeuGluLeuAlaGluAsnArgGluIleLysLysGluProValHisGlyValTyrTyrAsp 899  
1621 CCCAGCAAGACCTGTGGCGGAGATCCAGAGCGGCGACACCGACCTACCG 1680  
900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
1681 ATCTACAGGAGCGCTTCAAGAACTCAAGACCGGCAAGTACGCAAGATCGCGACCGCC 1740  
920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
1741 CACACCAACGATGAGCAGCTGACCGAGCGCTGAGAGATCGCCATCGGAGATC 1800  
940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959  
1801 GTGATCTGGGCGAAGACCCCAAGTCTCGCTGCCCATCCAGAGGAGACCTGGGAGAC 1860  
960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
1861 TGTGGACCGACTACTGGCAGGCGCCACCTGGATCCCGGAGTGGAGTTCGTGAACACCC 1920  
980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
1921 CCCTGTGTAGTGTGTGTTACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980  
1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

QY 1981 TACGTGACGCGCGCCGCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2040  
Db TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039  
QY 2041 CGGGCGCGGAGAAGATCGTGAGCTGACCGGAGACCAACCAAGACCGAGCTGCAG 2100  
Db ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
QY 2101 GCCATCCAGCTGGCGCTCGAGGACGAGCGAGGAGTGAACATCGTGACCGACGAC 2160  
Db AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079  
QY 2161 TACGCTCGGCGCATCTCCAGCGCCCGACCAAGAGCGAGCGAGCTGGTGAACCGAC 2220  
Db TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
QY 2221 ATCATCCAGCAGCTGATCAAGAGGAGAGTGTACTCTGAGCTGGTGGCGCCCGAC 2280  
Db IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
QY 2281 GGCAATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGCATCCGCAAGTGTG 2340  
Db GlyIleGlyGlnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
QY 2341 TTCCTGCGCGCATCGAT 2358  
Db PheLeuAspGlyIleAsp 1145  
RESULT 8  
Q6X4R4 PRELIMINARY; PRT; 1430 AA.  
AC Q6X4R4  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Gag-pol fusion polypeptide (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14709249; DOI=10.1089/08922203771881220;  
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,  
RA McCutchan F.E.;  
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
in full genome analysis, is not sustained in subgenomic regions.";  
RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Harris M.E., Birx D.L.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kim B., McCutchan F.E.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Maayan S., Zeira M.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AY255823; AAP76521.1; --  
DR HSP; P12497; I89D  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.



Qy	1369	GACAAGTGAACCGTGCAGCCCATCGACTGCGCGGAGAAAGAGAGACTGGACCGTGAACGAC	1428
Db	819	AspLysTrpThrValGlnThrIleGlnLeuProGlnLysAspSerTrpThrValAsnAsp	838
Qy	1429	ATCCAGAAAGCTGGTGGCGCAAGCTCAACTGGGCGCAGCCAGATCTACCCCGCGCATCAAGGTG	1488
Db	839	IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal	858
Qy	1489	CGCCAGCTGTGCAAGCTGCTGCGCGCGCCCAAGCCCTGACCGGACATCGTGCCTCCATGACC	1548
Db	859	ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValAlaLeuThr	878
Qy	1549	GAGAGAGCCGAGCTGGAGCTGGCCGCGAAGCCGCGAGATCTTGGCGGAGCCCGTGCACGGC	1608
Db	879	GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGly	898
Qy	1609	GTGTACTACGACCCACGACGAGCTGTGTGCGCGAGATCCAGAGCAGGAGGCCACGACCGAG	1668
Db	899	ValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln	918
Qy	1669	TGGACCTTACCAGATCTTACCAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTAGTACGCCAAG	1728
Db	919	TrpThrTyrGlnIleTyrGlnLeuProPheLysAsnLeuLysThrGlyLysTyrAlaLys	938
Qy	1729	ATGGCGACCGCCACACAAACGAGCTGAAGCAGTGAACGAGGCCGTGCAGAGATGCGC	1788
Db	939	ArgArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla	958
Qy	1789	ATGAGAGCATCTGATCTGGGCAACACCCCAAGTTCGCGCTGCCATCCAGAGTGGGAGTTC	1848
Db	959	MetGluSerIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGlu	978
Qy	1849	ACCTGGGAGACCTGGTGGACCGCACTACTGGCAGCGCACCTCGATCCCGAGTGGGAGTTC	1908
Db	979	ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe	998
Qy	1909	GTGAACACCCCCCTGGTGAAGCTGTGTGATCCAGCTGGAGAGAGAGCCCATCATCGCC	1968
Db	999	ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly	1018
Qy	1969	GCCGAGACCTTCACTGCTGGAGCGCGCGCCCAACCGCGAGACCAAGATCCGCAAGCGCGC	2028
Db	1019	AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly	1038
Qy	2029	TACGTGACCAACCGCGCGCGCAGAACTCGTGAAGCTGTGACCGGAGACCAACCAAGAG	2088
Db	1039	TyrValThrAspLysGlyArgLysValValSerLeuAsnGluThrThrAsnGlnLys	1058
Qy	2089	ACCGAGCTGCAGGCCATCCAGCTGGCCCTCGAGGACGCGGACGAGTGAACATCGTG	2148
Db	1059	ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal	1078
Qy	2149	ACCGACAGCCAGTACCGCCCTGGGATCATTCAGGCCCGCGCCCAACAGACGAGAGCGAG	2208
Db	1079	ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerAspSerGlu	1098
Qy	2209	CTGTGTGAACAGATCATCGAGCAGCTCATCAAGAAGGAGAGGTGTACCTGAGCTGGGTG	2268
Db	1099	IleValAsnGlnIleIleGlnIleLysGluArgValTyrLeuSerTrpVal	1118
Qy	2269	CCCGCCCAAGGGCATCGCGCGCAACGAGCAGATCGACAACTGGTGTGACCAAGGGCATC	2328
Db	1119	ProAlaHisLysGlyIleGlyGlyAsnGlnValAspLysLeuValSerLysGlyIle	1138
Qy	2329	CGCAAGTGTCTTCCTGGACGGCATCGAT	2358
Db	1139	ArgLysValLeuPheLeuAspGlyIleAsp	1148
RESULT 9			
Q9WF71			
ID	Q9WF71	PRELIMINARY;	PRT; 1427 AA.
AC	Q9WF71;		
DT	01-NOV-1999	(TrEMBLrel. 12, Created)	

01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
Gag-pol polyprotein.  
Human immunodeficiency virus 1.  
Viruses: Retroid viruses; Retroviridae; Lentivirus.  
NCBI\_TaxID=11676;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=92214383; PubMed=10196340;  
Novitsky V.A., Montano M.A., McLane M.F., Vannberg F.,  
Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
Essex M.;  
"Molecular cloning and phylogenetic analysis of human immunodeficiency  
virus type 1 subtype C: a set of 23 full-length clones from  
Botswana.";  
J. Virol. 73:4427-4432(1999).  
-1- SIMILARITY: Belongs to peptidase family A2.  
EMBL; AF110964; AAD17047.1; -;  
HSSRP; Q70622; IHVN.  
GO; GO:0019012; C:virion; IEA.  
GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
GO; GO:0003677; F:DNA binding; IEA.  
GO; GO:0008907; F:integrase activity; IEA.  
GO; GO:0008233; F:peptidase activity; IEA.  
GO; GO:0004523; F:ribonuclease H activity; IEA.  
GO; GO:0003723; F:RNA binding; IEA.  
GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
GO; GO:0005198; F:structural molecule activity; IEA.  
GO; GO:0016740; F:transferase activity; IEA.  
GO; GO:0008270; F:zinc ion binding; IEA.  
GO; GO:0015074; P:DNA integration; IEA.  
GO; GO:0006310; P:DNA recombination; IEA.  
GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
GO; GO:0016032; P:viral life cycle; IEA.  
InterPro; IPR000721; Gag\_p24.  
InterPro; IPR000721; Integrase C.  
InterPro; IPR003308; Integrase Zn N.  
InterPro; IPR001995; Peptidase A2.  
InterPro; IPR009007; Pept Aspartic.  
InterPro; IPR001969; Pept\_Asp\_AS.  
InterPro; IPR010999; Retrovir\_matrix.  
InterPro; IPR000071; Retrovir\_p17.  
InterPro; IPR008916; Retrov capsid C.  
InterPro; IPR008919; Retrov capsid N.  
InterPro; IPR002156; RNaseH.  
InterPro; IPR001584; Rve.  
InterPro; IPR000477; RVtse.  
InterPro; IPR010659; RVT connect.  
InterPro; IPR010661; RVT thumb.  
InterPro; IPR001878; Znf\_CCHC.  
Pfam; PF00540; Gag\_p17; 1.  
Pfam; PF06007; Gag\_p24; 1.  
Pfam; PF2022; Integrase\_Zn; 1.  
Pfam; PF00075; RNaseH; 1.  
Pfam; PF00665; rve; 1.  
Pfam; PF00077; RVP; 1.  
Pfam; PF00078; RVT\_1; 1.  
Pfam; PF08815; RVT\_connect; 1.  
Pfam; PF08817; RVT\_thumb; 1.  
Pfam; PF00098; zF\_CCHC; 2.  
PRINTS; PR00939; C2HCZNFINGER.  
PRINTS; PR00234; HIVIMATRIX.  
SMART; SM00343; Znf\_C2HC; 2.  
PROSITE; PS00141; ASP\_PROTEASE; 1.  
PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
PROSITE; PS0158; ZF\_CCHC; 2.  
AIDS; Aspartyl protease; Core protein; Hydrolyase; Metal-binding;  
polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
zinc-finger.  
SEQUENCE 1427 AA; 161326 MW; B9FC43C29B26CFA2 CRC64;  
Alignment Scores:

Alignment Scores:



```

Db 1040 LysGlyArgGlnLysIleValSerLeuAsnGluThrThrAsnGlnLysAlaGluLeuGln 1059
Qy 2101 GCCATCCAGCTGGCCCTCAGACAGCGGACAGGAGGTAACATCGTGACCGACGAGCCAG 2160
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079
Qy 2161 TACGCCCTGGGCATCATCCAGCCCGACCCGACAGAGCGAGCGAGCTGGTGAACCCAG 2220
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099
Qy 2221 ATCATCGAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGTGGCCGCCACAG 2280
Db 1100 IleIleGluGlnLeuIleLysGlyLeuArgValTyrLeuSerTrpValProAlaHisLys 1119
Qy 2281 GGCAATCGCGCGCAACAGCAGCATGCACAAGCTGGTGAGCAAGGCGATCCGCAAGTGCTG 2340
Db 1120 GlyIleGlyGlnGlnGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139
Qy 2341 TTCCTGAGCGCATCGAT 2358
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 10
Q9WFA0 PRELIMINARY; PRT; 1427 AA.
AC Q9WFA0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vamberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RL Botswana."
RL J. Virol. 73:4427-4432(1999).
CC -!- SIMILARITY: Belongs to peptidase family A2.
DR EMBL; AF110978; AAD17162.1, -.
DR PIR; S49086; S49086.
DR HSSP; Q70622; IHVN.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0015074; F:DNA integration; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR001037; Integrase C.
DR InterPro; IPR003308; Integrase Zn N.
DR InterPro; IPR001995; Peptidase A2.
DR InterPro; IPR009007; Pept. Aspartic.
DR InterPro; IPR001969; Pept. Asp. AS.
DR InterPro; IPR010999; Retrovir. matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.

```

```

DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR010659; RVT_connect.
DR InterPro; IPR010661; RVT_chumb.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; Rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT_1; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF06817; RVT_chumb; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C3HCZNFINGER.
DR PRINTS; PR00234; HIVIMATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
DR PROSITE; PS00158; ZF_CCHC_2.
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1427 AA; 160503 MW; 24C410797550C0F8 CRC64;

Alignment Scores:
Pred. No.: 7,37e-147 Length: 1427
Score: 3888.00 Matches: 723
Percent Similarity: 96.44% Conservative: 35
Best Local Similarity: 91.98% Mismatches: 24
Query Match: 85.36% Indels: 4
DB: 2

US-09-610-313B-32 (1-2457) x Q9WFA0 (1-1427)
Qy 13 ATGCGCGAGCGCATGAGCCAGGCGCCAGCCAGCCCAACATCTGTGATGAGCGAGCAACTTC 72
Db 360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379
Qy 73 AAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGGAGGCGCCATCGCCGCG 132
Db 380 LysGlyProArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399
Qy 133 AACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 192
Db 400 AsnCysArgAlaProArgLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419
Qy 193 AAGGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252
Db 420 LysAspCysThrGluArgGlnAlaAspPheArgGluAsnLeuAlaPheProGlnGly 439
Qy 253 AAGCGCGCGCGAGTTCCTCCAGCGAGCAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 312
Db 440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerGlyGluLeu 459
Qy 313 CAGTGGCGCGCGCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db 460 GlnValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479
Qy 373 TTTCCCGCAGATCACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnVal 499
Qy 433 AAGGAGCGCGCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspThrValLeuGluGluIleAsnLeuPro 519
Qy 493 GGCAAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 552
Db 520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyr 539

```



QY	553	GACCAGATCTCTGATCGAGATCTGGCGAAGAAGGCCATCGGCACCGCTGATCGGCCCC	612
DB	540	AspGlnIleThrIleGluIleCysGlyIleValIleGlyThrValLeuValGlyPro	559
QY	613	ACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGCGTGCACCCCTGAACCTTC	672
DB	560	ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe	579
QY	673	CCCATCAGCCCATCGAGACCGTCCCGCTGAAGCTGAAGCCCGGCATGGACGGCCCCCAAG	732
DB	580	ProIleSerProIleGluThrValProValIleLeuLysProGlyMetAspGlyProLys	599
QY	733	GTGAAGCAGTGGCCCTTACCGAGGAGAAGATCAAGGCCCTGTACCGCCATCTCTCGAGGAG	792
DB	600	ValIysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu	619
QY	793	ATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACACACCCCGCTG	852
DB	620	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal	639
QY	853	TTCCGCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGTGTGACTTCCCGCAGCTG	912
DB	640	PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
QY	913	AACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCATGCCCCACCCGCGCGGCTG	972
DB	660	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	679
QY	973	AAGAAGAGAGACGCTGACGCTGCTGAGCTGGCGCAGCCCTACTTACGCGTGGCCCTG	1032
DB	680	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
QY	1033	GACGAGGACTTCCGCAAGTACACCGCCTTCCACATCCCCAGCATCAACACGAGACCCCC	1092
DB	700	AspGluSerPheAspGlySerThrAlaPheThrIleProSerIleAsnAsnAlaThrPro	719
QY	1093	GGCATCGCTACAGTACAGTGTCTGCTGCCAGGGCTGGGAAGGCGACCCCGACGATCTTC	1152
DB	720	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	739
QY	1153	CAGACGACATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGTATC	1212
DB	740	GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle	759
QY	1213	TACCAG-----GCCCCCTGTACGTGGGCAGCGACTGTGAGATCGGCCACGACCGGCC	1266
DB	760	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
QY	1267	AAGATCGGAGCTGCCAAGCACCTCTCGCTGGGGCTTCACCAACCCCGCAGACAGAG	1326
DB	780	LysIleGluLeuLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys	799
QY	1327	CACCAGAAGAGGCCCTTCCTCTG-----CCCATCGAGCTGCACCCCGCAACAGTGGACC	1380
DB	800	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	819
QY	1381	GTGCAGGCCCATCGAGTGCCTCCGAGAAGGAGAGCTGGACCTGTGACGACATCCAGAAGCTG	1440
DB	820	ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu	839
QY	1441	GTGGGCAAGCTGAACCTGGCCAGCCAGATATACCCCGCATCAAGTGGCCGAGCTGTC	1500
DB	840	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys	859
QY	1501	AAGCTGTGTCGGCGCGCAAGGCCCTGCACCGATCTGTCGCCCTGCACCGAGGAGCGCGAG	1560
DB	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu	879
QY	1561	CTGAGCTGGCCGAGAACCGCGAGATCTCTGGCGAGCCCGTGCACGCGTGTTACTAGCAC	1620
DB	880	LeuGluLeuAlaGluAsnArgGluLeuLysGluSerProValHisGlyValTyrTyrAsp	899

```

RESULT 11
Q9WF65
ID Q9WF65 PRELIMINARY; PRT; 1427 AA.
AC Q9WF65;
AD 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Vannberg F.,
RA Foley B.T., Nung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.;
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana.";
RL J. Virol. 73:4427-4432(1999) .

```

-1- SIMILARITY: Belongs to peptidase family A2.  
EMBL; AF110962; AAD17030.1; -.

HSPB; Q70622; 1HVN.

DR GO:0019012; C:varion; IEA.

DR GO:0004190; F:aspartic-type endopeptidase activity; IEA.

DR GO:0003677; F:DNA binding; IEA.

DR GO:0008907; F:peptidase activity; IEA.

DR GO:0008233; F:peptidase activity; IEA.

DR GO:0004523; F:ribonuclease H activity; IEA.

DR GO:0003723; F:RNA binding; IEA.

DR GO:0003964; F:RNA-directed DNA polymerase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.

DR GO:0016740; F:transferase activity; IEA.

DR GO:0008270; F:zinc ion binding; IEA.

DR GO:0015074; P:DNA integration; IEA.

DR GO:0006310; P:DNA recombination; IEA.

DR GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO:0006278; P:RNA-dependent DNA replication; IEA.

DR GO:0016032; P:viral life cycle; IEA.

DR InterPro: IPR000721; Gag\_p24.

DR InterPro: IPR001037; Integrase C.

DR InterPro: IPR003308; Integrase Zn N.

DR InterPro: IPR001995; Peptidase A2.

DR InterPro: IPR009007; Pept. Aspartic.

DR InterPro: IPR001969; Pept. Asp. AS.

DR InterPro: IPR010999; Retrovir. matrix.

DR InterPro: IPR000071; Retrovir. p17.

DR InterPro: IPR008916; Retrov. capsid C.

DR InterPro: IPR008919; Retrov. capsid N.

DR InterPro: IPR002156; RNaseH.

DR InterPro: IPR001584; Rve.

DR InterPro: IPR000477; RVTse.

DR InterPro: IPR010659; RVT connect.

DR InterPro: IPR010661; RVT thumb.

DR InterPro: IPR001878; Znf\_CCHC.

DR Pfam: PF00540; Gag\_p17; 1.

DR Pfam: PF00607; Gag\_p24; 1.

DR Pfam: PF02022; Integrase\_Zn; 1.

DR Pfam: PF00075; RNaseH; 1.

DR Pfam: PF00665; rve; 1.

DR Pfam: PF00077; RVP; 1.

DR Pfam: PF00078; RVT\_1; 1.

DR Pfam: PF06815; RVT\_connect; 1.

DR Pfam: PF06817; RVT\_thumb; 1.

DR Pfam: PF00098; zf\_CCHC; 2.

DR SMART: SM00343; Znf\_C2HC; 2.

DR PROSITE: PS00141; ASP\_PROTEASE; 1.

DR PROSITE: PS01175; ASP\_PROT\_RETROV; 1.

DR PROSITE: PS0158; ZF\_CCHC; 2.

KW AIDS; Aspartyl protease; Core protein; Hydrolase; Polypeptide;

KW Protease; RNA-directed DNA polymerase; Transferrase.

SQ SEQUENCE 1427 AA; 161347 MW; 595F19865DE81D63 CRC64;

#### Alignment Scores:

Pred. No.:	9.26e-147	Length:	1427
Score:	3885.50	Matches:	724
Percent Similarity:	96.44%	Conservative:	34
Best Local Similarity:	92.11%	Mismatches:	23
Query Match:	85.30%	Indels:	5
DB:	2	Gaps:	3

US-09-610-313B-32 (1-2457) x Q9WF65 (1-1427)

QY 13 ATGCGCGCGCATGAGCCAGCCAGCCCAACATCTGATGCGCGCAACTTC 72

Db 361 LeuAlaGluLaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe 379

QY 73 AAGGCGCCCAAGCCCATCATCAAGTGTTCACATGCGCGCAAGAGGCGCCACATCGCCCGC 132

Db 380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

QY 133 AACTGCCGCGCCCGCCCAAGAGGGTGTGGAAGTGGCGCAAGAGGGCCACCATG 192

Db 133 AACTGCCGCGCCCGCCCGCCCAAGAGGGTGTGGAAGTGGCGCAAGAGGGCCACCATG 192

QY 133 AACTGCCGCGCCCGCCCGCCCGCCCAAGAGGGTGTGGAAGTGGCGCAAGAGGGCCACCATG 192

Db 133 AACTGCCGCGCCCGCCCGCCCGCCCGCCCAAGAGGGTGTGGAAGTGGCGCAAGAGGGCCACCATG 192

QY 133 AACTGCCGCGCCCGCCCGCCCGCCCGCCCGCCCAAGAGGGTGTGGAAGTGGCGCAAGAGGGCCACCATG 192

Db	400	Asn	Cys	Arg	Ala	Pro	Arg	Lys	Gly	Cys	Trp	Lys	Cys	Gly	Lys	Glu	Gly	His	Gln	Met	419
QY	193	AAG	ACT	GC	AC	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	252
Db	420	Lys	Asp	Cys	Thr	Glu	Arg	Gln	Ala	Asn	Phe	Arg	Glu	Asn	Leu	Ala	Phe	Pro	Gln	Gly	439
QY	253	AAG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	312
Db	440	Glu	Ala	Arg	Glu	Phe	Pro	Ser	Glu	Gln	Thr	Arg	Ala	Asn	Ser	Pro	Thr	Ser	Arg	Glu	459
QY	313	CAG	TG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	372
Db	460	Gln	Val	Arg	Gly	Asp	Asn	Pro	Arg	Ser	Glu	Thr	Arg	Ala	Glu	Gly	Gln	Gly	Thr	Leu	479
QY	373	TT	CC	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	432
Db	480	Phe	Pro	Gln	Ile	Thr	Leu	Trp	Gln	Arg	Pro	Leu	Val	Ser	Ile	Val	Gly	Gly	Gln	Ile	499
QY	433	AAG	AG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	492
Db	500	Arg	Glu	Ala	Leu	Leu	Asp	Thr	Gly	Ala	Asp	Thr	Val	Leu	Glu	Asp	Ile	Asn	Leu	Pro	519
QY	493	GG	CA	AG	TG	GA	AG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	552
Db	520	Gly	Lys	Trp	Lys	Pro	Lys	Met	Ile	Gly	Gly	Ile	Gly	Gly	Phe	Ile	Lys	Val	Arg	Gln	539
QY	553	GAC	CAG	AT	CT	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	612
Db	540	Glu	Gln	Ile	Leu	Ile	Glu	Ile	Cys	Gly	Lys	Lys	Ala	Ile	Gly	Thr	Val	Leu	Val	Gly	559
QY	613	ACC	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	672
Db	560	Thr	Pro	Val	Asn	Ile	Gly	Arg	Asn	Met	Leu	Thr	Gln	Leu	Gly	Cys	Thr	Leu	Asn	Phe	579
QY	673	CC	AT	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	732
Db	580	Pro	Ile	Ser	Pro	Ile	Glu	Thr	Val	Pro	Val	Lys	Leu	Lys	Pro	Gly	Met	Asp	Gly	Pro	599
QY	733	GT	GA	AG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	792
Db	600	Val	Lys	Gln	Trp	Pro	Leu	Thr	Glu	Glu	Lys	Ala	Leu	Thr	Ala	Ile	Cys	Glu	Glu	619	
QY	793	AT	GG	AG	AG	GG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	852
Db	620	Met	Glu	Lys	Glu	Gly	Lys	Ile	Thr	Lys	Ile	Gly	Pro	Glu	Asn	Pro	Trp	Asn	Thr	Pro	639
QY	853	TT	GC	CG	AT	CA	AG	AG	AG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	912
Db	640	Phe	Ala	Ile	Lys	Lys	Lys	Asp	Ser	Thr	Lys	Trp	Arg	Lys	Leu	Val	Asp	Phe	Arg	Glu	659
QY	913	AAC	AA	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	972
Db	660	Asn	Lys	Arg	Thr	Gln	Asp	Phe	Trp	Glu	Val	Gln	Leu	Gly	Ile	Pro	His	Pro	Ala	Gly	679
QY	973	AAG	AA	AG	AG	AG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	1032
Db	680	Lys	Lys	Lys	Lys	Ser	Val	Thr	Val	Leu	Asp	Val	Gly	Asp	Ala	Val	Phe	Ser	Val	Pro	699
QY	1033	GAC	GAG	ACT	TC	CG	CA	AG	TAC	AC	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	1092
Db	700	Asp	Glu	Ser	Phe	Arg	Lys	Trp	Thr	Ala	Phe	Thr	Pro	Ser	Ile	Asn	Asn	Ser	Thr	Pro	719
QY	1093	GG	AT	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	1152
Db	720	Gly	Lys	Arg	Trp	Gln	Trp	Asn	Val	Leu	Pro	Gln	Gly	Trp	Lys	Gly	Ser	Pro	Ala	Ile	739
QY	1153	CAG	AG	CG	AT	GC	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	1212
Db	740	Gln	Ser	Ser	Met	Ile	Lys	Ile	Leu	Glu	Pro	Phe	Arg	Thr	Lys	Asn	Pro	Asp	Ile	Val	759
QY	1213	TAC	CAG	-----	GC	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	1266
Db	760	Tyr	Gln	Trp	Met	Asp	Asp	Leu	Trp	Val	Gly	Ser	Asp	Leu	Glu	Ile	Gly	Gln	His	Arg	779

QY 1267 AAGATCGAGAGCTGGCAAGCACCTGCTGGGCTTCCACCAACCCCGACCAAG 1326  
 DB 780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProTyrLysLys 799  
 QY 1327 CACCAAGAGAGCCCTTCCTG-----CCATCGAGCTGCACCCCGACAAGTGGACC 1380  
 DB 800 HisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
 QY 1381 GTGACGCCATCGAGCTGCCAGAGAGAGAGCTGACCGTGAACGACATCCAGAGCTG 1440  
 DB 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValHisAspIleGlnLysLeu 839  
 QY 1441 GTGGCAAGCTCAACTGGGCCAGCAGATCTACCCCGGCATCAAGGTGGCCAGCTGCC 1500  
 DB 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTrpProGlyIleArgValLysHisLeuCys 859  
 QY 1501 AAGCTGCTCGCGGGCCCAAGGCCCTTGACCGATCTGTCGCCCTGACCGAGAGGCCGAG 1560  
 DB 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
 QY 1561 CTGAGCTGGCCGAGAACCGCAGAGATCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620  
 DB 880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp 899  
 QY 1621 CCACGCAAGGACTGTGTGGCCGAGATCCAGAGAGCGGCCACGACGAGTGGACTTACCAG 1680  
 DB 900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
 QY 1681 ATCTACCGAGGCCCTTCAAGAACTCTGAAGACCGGCAAGTACGCCAAGATGCGCCGCC 1740  
 DB 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetargThrAla 939  
 QY 1741 CACACCAACGAGTGAAGCAGCTGACCGAGCGCGTGCAGAAATCGCCATCGAGAGCATC 1800  
 DB 940 HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleThrMetGluSerIle 959  
 QY 1801 GTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCTCCACAGAGAGACCTGGGAGACC 1860  
 DB 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979  
 QY 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCC 1920  
 DB 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
 QY 1921 CCCCTGGTGAAGCTGTGTGATCAGCTGGAGAGAGAGCCCATCATCGCGCCGAGACCTTC 1980  
 DB 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019  
 QY 1981 TACGTGGACGGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2040  
 DB 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039  
 QY 2041 CGGGCGCGCAGAGATCTGAGCTGACCGAGACCCAGACCAACCAAGAGACCCAGCTGCAG 2100  
 DB 1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnArgAlaGluLeuGln 1059  
 QY 2101 GCATCCAGCTGGCCCTCGAGACAGCGCAGCGAGGTGAACATCGTGAACCAAGCCGAG 2160  
 DB 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
 QY 2161 TACGCTGGGATCATCAGCGCCAGCCCGACACAGCGAGCGAGCGTGGTGAACAG 2220  
 DB 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099  
 QY 2221 ATCATCGACGAGCTGATCAAGAGAGAGAGTGTACTCGAGCTGGGTGCCCGCCCAAG 2280  
 DB 1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
 QY 2281 GGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGGAGCAAGGCGATCCGCAAGTGTCTG 2340  
 DB 1120 GlyIleGlyGlyAsnGluValAspLysLysLeuValSerSerGlyIleArgLysValLeu 1139

QY 2341 TTCTGCGAGCGCATCGAT 2358  
 DB 1140 PheLeuAspGlyIleAsp 1145  
 RESULT 12  
 Q9WF77  
 ID Q9WF77 PRELIMINARY; PRT; 1437 AA.  
 AC Q9WF77;  
 DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DT 01-NOV-2004 (TRENBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
 RA Essex M.,  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana.",  
 RL J. Virol. 73:4427-4432(1999).  
 CC 1- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AF110967; AAD17072.1; -.  
 DR HSSP; P24736; INCP.  
 DR GO; GO:0019012; C:virion; IEA.  
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008907; F:integrase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0005198; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0003504; F:DNA integration; IEA.  
 DR GO; GO:0006310; F:DNA recombination; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn\_N.  
 DR InterPro; IPR001995; Peptidase\_A2.  
 DR InterPro; IPR009007; Pept\_Aspartic.  
 DR InterPro; IPR001969; Pept\_Asp\_AS.  
 DR InterPro; IPR010999; Retrovir\_matrix.  
 DR InterPro; IPR000071; Retrovir\_p17.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVise.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF06815; RVT\_connect; 1.  
 DR Pfam; PF06817; RVT\_thumb; 1.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PRINTS; PR00234; HIVMATRIX.  
 DR SMART; SM00343; Znf\_C2HC; 2.





Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnIle 499  
Qy 433 AAGGAGGCGCTGCTGGAGACCGCGGCGCCGACACACCGCTGCTGGAGAGATGAGCTGCC 492  
Db 500 ArgGluAlaLeuLeuAspThrGlyAlaAspThrValLeuGluAspIleAsnLeuPro 519  
Qy 493 GGCAGTGGAGCAAGATGATCGGCGGCATCGCGGCTTCATCAAGTGGCGCAGTAC 552  
Db 520 GlyLysTrpLysProLysMetIleGlyGlyPheIleLysValArgGlnTyr 539  
Qy 553 GACCAGATCCTGATCGAGATCTGGGCAAGAGCCATCGCACCGTCTGATCGGCGCC 612  
Db 540 GluGlnIleLeuIleGluIleCysGlyLysAlaIleGlyThrValLeuValGlyPro 559  
Qy 613 ACCCGCTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTCAACTTC 672  
Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe 579  
Qy 673 CCATACAGCCCATCGAGACCGTCCGCTGAAGCTGAAGCCCGGATGAGCGGCCCAAG 732  
Db 580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
Qy 733 GTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGCCCTGACCGCATCTCGAGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619  
Qy 793 ATGAGAGAGGAGGCAAGATCACAAGATCGCGCCCGAGAACCCCTACAAACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
Qy 853 TTCGCCATCAAGAAAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAGGCGCACCCAGGACTTCTGGAGTGTGGAGTGGCATCCCGACCCCGCGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAGAGAGCGTGCCTGCTGACGTGGCGACGCTACTTTCAGCGTCCCGCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGGACTTCCCAAGTACACCGCTTCCACCATCCCGAGCATCAACAAGACACCC 1092  
Db 700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719  
Qy 1093 GGCATCGCTACAGTACAGTCTGCTGCCCGAGGCTGGAGGCGACCCCGAGATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
Qy 1153 CAGAGCAGCATGACCAAGATCTCGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759  
Qy 1213 TACCAG-----GCCCGCTGTAGCTGGGCGACCTTCCGCGCCGCAACCCCGAGATCGTGATC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1267 AAGATCAGGAGCTGCGCAGCAGCATCTGCTGGCTGGGCTTCCACACCCCGCAGCAAG 1326  
Db 780 LysIleGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrProAspLysLys 799  
Qy 1327 CACCAGAGGAGCGCCCTTCCTG-----CCCATCGAGTGCACCCCGACAGTGAGCC 1380  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1381 GTGAGCGCATCGAGCTGCCGAGAGCAGAGCTGGACCGTGAACGACATCCAGAGCTG 1440  
Db 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1441 GTGGCGAGCTGAATCGGCGCAGCAGATCTACCCCGGATCAAGGTCGCGCAGCTGTC 1500  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValArgHisLeuCys 859

Qy 1501 AAGTGTCTGGCGCGCCAAAGCGCCTGACCGACATCGTGGCCCTGACCGAGGAGCGCAG 1560  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu 879  
Qy 1561 CTGGAGCTGGCGGAGAACCGCGAGATCTCTGGCGCGCCCGTGCACGGGCTGTACTACGAC 1620  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp 899  
Qy 1621 CCAGCAAGACCTGGTGGCGCGAGATCCAGAAGCAGGCGCACGACCTAGTGACCTACCAG 1680  
Db 900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1681 ATCTACCGAGGAGCCTTCAAGAACCTGAAGACCGCGCAAGTACGCAAGATGCGCACCGCC 1740  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1741 CACACCAACAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGGCATCGGAGCATC 1800  
Db 940 HisThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle 959  
Qy 1801 GTGATCTGGGCGAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979  
Qy 1861 TGGTGAACCGACTACTGCGCAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCC 1920  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1921 CCCTGTGTGAAGCTGTGTACCTGAGCTGGAGAGAGCCCATCATCTGGCGCGCAGACTTC 1980  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019  
Qy 1981 TAGCTGACCGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTAGCTACCGAC 2040  
Db 1020 TyrValAspGlyAlaAlaAsnArgLysLeuGlyLysAlaGlyTyrValThrAsn 1039  
Qy 2041 CGGCGCGCGCAGAGATCTGTGAGCTGACCGCAGACCAACCAAGAGACCGAGCTGCAG 2100  
Db 1040 LysGlyArgGlnLysIleValProLeuThrGluThrAsnGlnLysAlaGluLeuGln 1059  
Qy 2101 GCATTCAGCTGGCGCTGCGAGCAGCGCAGCGAGGTGAACATCTGTGACCGCAGCGCAG 2160  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy 2161 TAGCCCTGGGCATCATCCAGGCGCCCGCAGACCGCAGACCGAGCGGCTGGTGAACCG 2220  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099  
Qy 2221 ATCATTCAGCAGCTGTATCAAGAAGGAGAAGTGTACCTGAGCTGGTGGCGCGCCACAAG 2280  
Db 1100 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
Qy 2281 GGCATCGCGCGCAACGAGCAGATCGAACAGCTGGTGAAGGCGCATCCGCAAGGTGCTG 2340  
Db 1120 GlyIleGlyAsnGluValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
Qy 2341 TTCTGACCGCATCGAT 2358  
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 14  
Q8ADX5 PRELIMINARY; PRT; 1427 AA.  
AC Q8ADX5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol fusion polypeptide (Fragment).  
GN Name=gag-pol;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;





QY 793 ATGAGAGAGGAGGAGATCAACAGATCGGCGCCCGAGAACCCCTACACACACCCCGGTG 852  
 Db 620 MetGluysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIle 639  
 QY 853 TTCGCCATCAAGAAAGAGGAGCAGACCAAGTGGCGGAGCTGGGAGCTTCCTCCGAGCTG 912  
 Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
 QY 913 AACAAAGCCACCCAGGACTTCTGGAGGTGAGCTGGGCTATCCCGCCCGCGGCTG 972  
 Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
 QY 973 AAGAAGAAAGAGCGTGCCTGCTGAGCTGGGCGAGCGCTACTTACAGGTGCCCTG 1032  
 Db 680 LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
 QY 1033 GACGAGACTTCCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAGCAGACCC 1092  
 Db 700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
 QY 1093 GGCAATCCGCTACCAAGTACAGCTGCTGCCCGAGGCTGGAGGCGAGCCCGAGCATCTTC 1152  
 Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
 QY 1153 CAGAGCAGCATGACCAAGATCCTGAGCCCTTCCGCGCCGCGCAACCCCGAGATCGTGATC 1212  
 Db 740 GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759  
 QY 1213 TACCAG-----GCCCGCTGTAGTGGGCGAGCGACTGGAGATCGGCGCAGCCGCGCC 1266  
 Db 760 TyrGlnTyrMetAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779  
 QY 1267 AAGATCAGGAGAGCTGCGCAAGCAGCTGCTGGCTGGGCTTCCACACCCCGCAGCAAGAAG 1326  
 Db 780 LysIleGluGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
 QY 1327 CACCAAGAGAGCCCGCTTCCTG-----CCCATCAGCTGCACCCCGCAAGTGAAC 1380  
 Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
 QY 1381 GTGCAGCCATCAGAGTCCCGCAGAGGAGAGCTGGACCTGACGACATCCAGCATCCAGAGCTG 1440  
 Db 820 ValGlnProIleGlnLeuProGluLysAspSerTrpValAsnAspIleGlnLysLeu 839  
 QY 1441 GTGGCAGAGCTGAATCGGCGCAGCAGCATCTACCCCGCATCAAGTGGCGCAGCTGTGC 1500  
 Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValLysGlnLeuLys 859  
 QY 1501 AAGCTGTGCGCGCGCCAGGCGCTGACCGACATCGTGCCTGACCGAGGAGCGCGAG 1560  
 Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
 QY 1561 CTGAGCTGCGGAGACCGGAGATCTCGCGAGCCCGCTGCAGCGCGCTGACTAGAC 1620  
 Db 880 LeuGluLeuAlaGluAsnArgLysLeuLysGluProValHisGlyValTyrTyrAsp 899  
 QY 1621 CCACAGCAGCAGCTGTGGCGGAGATCCAGAGCAGGCGCCACCGACCATGAGCTACAC 1680  
 Db 900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGln 919  
 QY 1681 ATCTACAGGAGCCCTTCAAGAACTCAAGACCGGCAAGTACGCAAGATGCGCAGCCG 1740  
 Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939  
 QY 1741 CACACCAACGCTGAGCAGCTGACCGAGCGCTGCAGAGATCGCCATGAGAGATC 1800  
 Db 940 HisThrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959  
 QY 1801 GTGATCTGGGCGAGAGCCCGCAAGTCTCGCTGCCCATCAGAGGAGAGACTGGAGACC 1860  
 Db 960 IleIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluAla 979  
 QY 1861 TGGTGGACCGACTACTGGCAGGCGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCC 1920

Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
 QY 1921 CCCCTGTGTAAGCTGTGCTACCGAGTGAAGAGAGCCCATCATCGGCGCGAGACCTTC 1980  
 Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019  
 QY 1981 TAGCTGACGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2040  
 Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039  
 QY 2041 CGGGCGCGCAGAGATCTGTGAGCTGACCGCAGACCAACCAACCAAGACCGAGCTGCAG 2100  
 Db 1040 ArgGlyArgGlnLysIleIleSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
 QY 2101 GCATCCAGCTGGCGCTGCGAGCAGCGCAGCGAGGTGAACATCGTGACCGAGCGCAG 2160  
 Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValThrAspSerGln 1079  
 QY 2161 TAGCCCTGGGCTATCTCCAGGCGCCCGCAGACGAGCGAGCGAGCTGGTGAACCGAG 2220  
 Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099  
 QY 2221 ATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGTGGCGCCCGCAG 2280  
 Db 1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
 QY 2281 GGCAATCGGCGCAGCAGCAGATCGACAAGCTGTGTGAGCAAGGCGCATCCGCAAGTGTG 2340  
 Db 1120 GlyIleGlyGlnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
 QY 2341 TTCCTGGACGCGCATCGAT 2358  
 Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 15  
 Q9WF79 ID Q9WF79 PRELIMINARY; PRT; 1425 AA.  
 AC Q9WF79;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana.";  
 RL J. Virol. 73:4427-4432(1999).  
 CC -1- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AF110969; AAD17087.1; --  
 DR HSP; P24736; INCP.  
 DR GO; GO:0019012; C:virion; IEA.  
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008907; F:integrase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0015074; P:DNA integration; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO: GO:0016032; P:viral life cycle; IEA.  
 DR InterPro: IPR000721; Gag\_p24.  
 DR InterPro: IPR001037; Integrase\_C.  
 DR InterPro: IPR003308; Integrase\_Zn\_N.  
 DR InterPro: IPR001995; Peptidase\_A2.  
 DR InterPro: IPR009007; Pept\_Aspartic.  
 DR InterPro: IPR001969; Pept\_Asp\_AS.  
 DR InterPro: IPR010999; Retrovir\_matrix.  
 DR InterPro: IPR000071; Retrovir\_p17.  
 DR InterPro: IPR008916; Retrov\_capsid\_C.  
 DR InterPro: IPR008919; Retrov\_capsid\_N.  
 DR InterPro: IPR002156; RNaseH.  
 DR InterPro: IPR001584; Rvs.  
 DR InterPro: IPR000477; RVse.  
 DR InterPro: IPR010659; RVT connect.  
 DR InterPro: IPR010661; RVT thumb.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00540; Gag\_p17; 1.  
 DR Pfam: PF00607; Gag\_p24; 1.  
 DR Pfam: PF02022; Integrase\_Zn; 1.  
 DR Pfam: PF00075; RNaseH; 1.  
 DR Pfam: PF00665; rve; 1.  
 DR Pfam: PF00077; RVP; 1.  
 DR Pfam: PF00078; RVT; 1.  
 DR Pfam: PF06815; RVT\_connect; 1.  
 DR Pfam: PF06817; RVT\_thumb; 1.  
 DR Pfam: PF00098; zf\_CCHC; 2.  
 DR PRINTS: PR00939; C2HCZNFINGER.  
 DR PRINTS: PR00234; HIV1MATRIX.  
 DR SMART: SM00343; Znf\_C2HC; 2.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE: PS00175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE: PS0158; ZF\_CCHC; 2.  
 KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
 KW Zinc-finger.  
 SQ SEQUENCE 1425 AA; 160569 MW; 6DC8A1ABF3C88DA7 CRC64;

## Alignment Scores:

Pred. No.: 7,95e-146 Length: 1425  
 Score: 3862.00 Matches: 723  
 Percent Similarity: 96.31% Conservative: 34  
 Best Local Similarity: 91.98% Mismatches: 23  
 Query Match: 84.79% Indels: 6  
 DB: 2 Gaps: 3

US-09-610-313B-32 (1-2457) x Q9WPF9 (1-1425)

QY 13 ATGGCGGAGCCATGAGCCGACCCAGCGCCACATCCTGATCGAGCGGAGCACTTC 72  
 DB 360 LeuAlaGluAlaMetSerGlnAlaAsnAsnSerAsnAlaMetMetGlnLysAsnAsnPhe 379  
 QY 73 AAGGCGCCCAAGCGCATCATCAAGTGCTTCAACTCGGCAAGGAGGCGCCACATCGCCGCGC 132  
 DB 380 LysGlyProThrArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
 QY 133 AACTCGCGCGCCCGCCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCCACCATG 192  
 DB 400 AsnCysLysAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
 QY 193 AAGGACTGACCGAGCGCCAGCGCAACTCTTCCGCGAGGACCTGGCTTCCCGCGGCGC 252  
 DB 420 LysAspCysThrGluArgArgAlaAsnPhePheArgGluAlaPheProGlnGly 439  
 QY 253 AAGGCGCGGAGTTCCTCCAGCGAGAGAACCCCGCCCAACAGCCCGGAGGAGGAGGAGG 312  
 DB 440 GluAlaArgGluPheProSerGluGlnThrGlyAlaAsnSerProThrSerArgLysLeu 459  
 QY 313 CAGGTGCGCGCGCAACACCCCGCAGCGAGGCGCGCGCGAGCGCGCCACCTCGAAC 372  
 DB 460 -----ArgGlyAsnAsnProCysSerGluAlaGlyAspGluArgGlnGlyThrLeuAsn 477

QY 373 TTCCCCCAGCATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCGCCAGATC 432  
 DB 478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnIle 497  
 QY 433 AAGGAGCCCTGTGTGGACACCGGCGCCGACACACCGTGTGGAGGAGATGAGCTGCCCC 492  
 DB 498 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517  
 QY 493 GGCAGTGGAGAGCCCAAGATGATCGGGGATCGGGGCTTCATCAAGGTGGCGCCAGTAC 552  
 DB 518 GlyLysTrpArgProLysMetIleGlyLysMetIleGlyLysPheIleLysValArgGlnTrp 537  
 QY 553 GACCAGATCCTGATCGAGATCTGGCGCAAGAGGCGCATCGGCACCGTCTCTGATCGGCCCC 612  
 DB 538 AspGlnIleProIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 557  
 QY 613 ACCCCCGTGAACATCATCGGCGCCCAACATGTCGCCACCGTGGGTGGCTGCACCTGAACCTC 672  
 DB 558 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPhe 577  
 QY 673 CCCATCAGCCCATCGAGACCGTGGCCCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAG 732  
 DB 578 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597  
 QY 733 GTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGCAGGAG 792  
 DB 598 ValLysGlnTrpProLeuThrGluLysLysAlaIleThrAlaIleCysGluGlu 617  
 QY 793 ATGGAGAGGAGGCGCAGATCACCAAGATCGGCCCGCCGAGAACCCCTACAAACACCCCGGTG 852  
 DB 618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTrpAsnThrProVal 637  
 QY 853 TTCCGCCATCAAGAAGAGGAGCAGACCAAGTGGCGGCAAGCTGGTGGATTCGCCGAGCTG 912  
 DB 638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657  
 QY 913 AACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGCATCCCCACCCCGCGCGCTG 972  
 DB 658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677  
 QY 973 AAGAAGAGAGAGCGTGACCGTGTGCACTGGCGGCGACGCTACTTACGCGTGCCTTC 1032  
 DB 678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeu 697  
 QY 1033 CAGCAGGACTTCCCGCAAGTACACGCTTCCACCATCCCCCAGCATCAACAGCAGACCCCC 1092  
 DB 698 AspGluSerPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717  
 QY 1093 GGCATCGCTACCATGATCAACGTGCTGCCCGAGGCTGGAAGGGCAGCGCCCGAGCATCTTC 1152  
 DB 718 GlyIleArgTrpGlnTrpAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 737  
 QY 1153 CAGCAGCAGTACCAAGATCCTGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212  
 DB 738 GlnSerSerMetThrLysIleLeuGluProPheArgAlaLysAsnProGluLeuValIle 757  
 QY 1213 TACCG-----GCCCGCTGTAGTGGGCGAGCGACTGGAGATCGGCGCAGCACCGCCG 1266  
 DB 758 TyrGlnTrpMetAspAspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgAla 777  
 QY 1267 AAGATCGAGGAGTGGCGCAAGCACCTGTGCGCTGGGGCTTCACCCACCCCGCAGCAAGAG 1326  
 DB 778 ArgIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797  
 QY 1327 CACCAGAGGAGCGCCCTTCTCTG-----CCCATCGAGCTGCACCCCGCAAGTGGAGC 1380  
 DB 798 HisGlnLysGluProProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThr 817  
 QY 1381 GTGCGCCCATCGAGCTCCCGAGAGGAGGAGTGGACGCTGACGACATCCAGAGCTG 1440  
 DB 818 ValGlnSerIleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 837  
 QY 1441 GTGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGCGCATCAAGGTGGCGCAGCTGTGC 1500

```
Db      838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys 857
Qy      1501 AAGCTGCTGCGCGCGCAAGCCCTGACGACATCGTCCCTGACCGAGGAGCGCGAG 1560
Db      858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProThrGluGluAlaGlu 877
Qy      1561 CTGAGCTGGCCGAGAACCGGAGATCTCGCGAGCCCGTGACCGCGTGTACTACGAC 1620
Db      878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp 897
Qy      1621 CCAGCAAGGACCTGCTGGCCGAGATCCAGAGAGGCGCACGACGAGTGGACCTACGAC 1680
Db      898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspGlnTrpThrTyrGln 917
Qy      1681 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCC 1740
Db      918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937
Qy      1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCATGGAGAGCATC 1800
Db      938 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle 957
Qy      1801 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAAGAGACCTGGGAGACC 1860
Db      958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla 977
Qy      1861 TGTGGACCGACTACTCGCAGGCGCACCTGGATCCCGAGTGGAGTTCGTGACACCCCC 1920
Db      978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 997
Qy      1921 CCCCTGTGAAGCTGTGTACTACGCTGGAGAGAGCCCATCATCGCGCGGAGACCTTC 1980
Db      998 ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe 1017
Qy      1981 TACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTAGTGACCGAC 2040
Db      1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037
Qy      2041 CGGGCGCGGAGAGATCGTGAGCGCTGACCGAGACCAACACCAAGAGACCGAGCTGCAG 2100
Db      1038 LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057
Qy      2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGAGCGAGGTGAACATCGTACCGCAGACGCGCAG 2160
Db      1058 AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln 1077
Qy      2161 TACGCCCTGGGCAATCATCCAGGCCCGACCGCAAGAGCGAGCGAGCTGGTGAACGAC 2220
Db      1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097
Qy      2221 ATCATCGAGCAGCTGATCAGAGAGAGAGTGTACTGCTGAGTGGTGGCCGCCACAG 2280
Db      1098 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1117
Qy      2281 GGCATCGGCGCAACGAGCAGATCGACAAGCTGTGTAGCAAGGGCATCCGCAAGGTGCTG 2340
Db      1118 GlyIleGlyAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu 1137
Qy      2341 TTCCTGACCGCATCGAT 2358
Db      1138 PheLeuAspGlyIleAsp 1143
```

Search completed: June 2, 2005, 05:05:45  
Job time : 305.158 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 04:23:54 ; Search time 152.794 Seconds  
(without alignments)

11117.427 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 4555

Sequence: 1 gtcagccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters: 2931222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlh  
-Q=/cgn2\_1/USFO\_spool/US09610313/runat\_31052005\_155138\_15193/app\_query.fasta\_1.7893  
-DB=Published Applications\_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09610313@cgn 1.1 550 @runat 31052005\_155138\_15193  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THRSADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pcp\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pcp\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pcp\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pcp\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pcp\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pcp\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	3527	77.4	998	16	US-10-332-413-4	Sequence 4, Appli
2	3480.5	76.4	850	16	US-10-332-413-10	Sequence 10, Appl
3	3454	75.8	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3432	75.3	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3409	74.8	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3409	74.8	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3387	74.4	1003	16	US-10-325-468-23	Sequence 23, Appl
8	3385	74.3	1003	16	US-10-325-468-35	Sequence 35, Appl
9	3384	74.3	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3384	74.3	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3376	74.1	995	15	US-10-296-734-2	Sequence 2, Appli
12	3362	73.8	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3208.5	70.4	1006	15	US-10-296-734-1471	Sequence 1471, Ap
14	3140.5	68.9	1014	14	US-10-301-661A-6	Sequence 6, Appli
15	3076.5	67.5	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3076.5	67.5	1350	16	US-10-380-641-35	Sequence 35, Appl
17	3076.5	67.5	1350	17	US-10-636-730-35	Sequence 35, Appl
18	2779.5	61.0	875	10	US-09-952-060-6	Sequence 6, Appli
19	2779.5	61.0	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2779.5	61.0	875	16	US-10-380-641-6	Sequence 6, Appli
21	2779.5	61.0	875	17	US-10-636-730-6	Sequence 6, Appli
22	2774	60.9	850	10	US-09-952-060-2	Sequence 2, Appli
23	2774	60.9	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2774	60.9	850	16	US-10-380-641-2	Sequence 2, Appli
25	2774	60.9	850	17	US-10-636-730-2	Sequence 2, Appli
26	2755.5	60.5	875	10	US-09-952-060-8	Sequence 8, Appli
27	2755.5	60.5	875	15	US-10-168-217A-8	Sequence 8, Appli
28	2755.5	60.5	875	16	US-10-380-641-8	Sequence 8, Appli
29	2755.5	60.5	875	17	US-10-636-730-8	Sequence 8, Appli
30	2750	60.4	850	10	US-09-952-060-4	Sequence 4, Appli
31	2750	60.4	850	15	US-10-168-217A-4	Sequence 4, Appli
32	2750	60.4	850	16	US-10-380-641-4	Sequence 4, Appli
33	2750	60.4	850	17	US-10-636-730-4	Sequence 4, Appli
34	2746	60.3	562	9	US-09-735-487-14	Sequence 14, Appl
35	2746	60.3	562	14	US-10-342-188-14	Sequence 14, Appl
36	2742	60.2	561	15	US-10-399-920-2	Sequence 2, Appli
37	2740	60.2	561	14	US-10-059-271-87	Sequence 87, Appl
38	2734	60.0	562	14	US-10-102-622-10	Sequence 10, Appl
39	2715	59.6	560	14	US-10-205-641-1	Sequence 1, Appli
40	2705.5	59.4	561	14	US-10-102-622-12	Sequence 12, Appl
41	2693	59.1	560	11	US-09-725-652-1	Sequence 29, Appl
42	2659	58.4	546	8	US-08-808-031A-29	Sequence 29, Appl
43	2646	58.1	979	14	US-10-271-181B-118	Sequence 118, App
44	2553	56.0	851	14	US-10-059-271-88	Sequence 88, Appl
45	2528	55.5	999	15	US-10-346-000A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-10-332-413-4  
; Sequence 4, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swanstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdool  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Their Genes nd Modifications and Derivatives T.  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924



; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 998  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-10-332-413-4

Alignment Scores:  
Pred. No.: 1,148-163 Length: 998  
Score: 3527.00 Matches: 662  
Percent Similarity: 96.51% Conservative: 29  
Best Local Similarity: 92.46% Mismatches: 21  
Query Match: 77.43% Indels: 4  
DB: 16 Gaps: 2

US-09-610-313B-32 (1-2457) x US-10-332-413-4 (1-998)

```
QY 223 TTCCGCGAGGACCTGGCTTCCCTCCCGAGGCGCCGCGAGTTCCTCCGAGCGAGCAAC 282
DB 1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20
QY 283 CCGCGCAACAGCCCAACAGCGCGAGCTGCAGGTGCGCGCGCAACACCCCGCGAGCGAG 342
DB 21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgAsnAsnProArgSerGlu 40
QY 343 GCCGCGCGCGAGCGCGAGCGCACCTGAACTTCCCGAGATCACCTGTGGAGCGCCCC 402
DB 41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60
QY 403 CTGGTGAGCATCAAGTGGCGCGCGAGATCAAGAGAGCGCCCTGCGACACCGCGCGCAC 462
DB 61 LeuValSerIleLysIleGlyGlnThrArgGluAlaLeuLeuAsnThrGlyAlaAsp 80
QY 463 GACACCGCTGCTGGAGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGC 522
DB 81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGly 100
QY 523 ATCGCGCGGCTTCATCAGGTGGCGCGAGTACGACCGAGATCTGATCGAGATCTCGCGAAG 582
DB 101 IleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyLys 120
QY 583 AAGGCCATCGGCACGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
DB 121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140
QY 643 CTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCATCAGCCCGCATCGAGACCGTCCCGTG 702
DB 141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160
QY 703 AAGCTGAAGCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 762
DB 161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180
QY 763 ATCAGGCGCTGACCGCATCTCGAGGAGATGGAAGAGCGCGCGCGCGCGCGCGCGCGCG 822
DB 181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluIleThrLysIle 200
QY 823 GCGCGCGAGAACCCCTACAACACCCCGCTGTCGCGCATCAAGAAGAAGACAGACCAAG 882
DB 201 GlyProGluAsnProTyAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220
QY 883 TGGCGCAAGCTGGTGGATTCCTCCGAGCTGAACAAAGCGCACCCAGGATCTTGGAGGTG 942
DB 221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240
QY 943 CAGCTGGGCATCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1002
DB 241 GlnLeuGlyIleProHieProAlaGlyLeuLysLysLysLysValThrValLeuAsp 260
QY 1003 GTGGCGGACCGCTTCTGAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTC 1062
```

```
DB 261 ValGlyAspAlaTyPheSerValProLeuAspGluGlyPheArgLysTyThrAlaPhe 280
QY 1063 ACCATCCCGCAGCATCAACAGAGACCCCGCGCATCCGCTACCACTAGTCAACAGTGTGCGCC 1122
DB 281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGlnTyAsnValLeuPro 300
QY 1123 CAGGCTGGAAGGCGCGCGCGCATCTTCAGAGCAGCATGACCAAGATCCTTGAGCGCC 1182
DB 301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320
QY 1183 TTCCGCGCGCGCAACCCCGAGATCGTGATCTACAG-----GCCCGCTGTAGCTGGCG 1236
DB 321 PheArgAlaGlnAsnProGluIleValIleTyGlnTyMetAspAspLeuTyValGly 340
QY 1237 AGCGACCTGGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1296
DB 341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360
QY 1297 CGCTGGGCTTCACACCCCGCACAGAGACACAGAGAGAGCGCCCTTCTCTG----- 1350
DB 361 LysTrpGlyPheThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380
QY 1351 CCCATCGAGCTGCACCCCGCACAGTGGACCGCTGCAGCCCATCGAGCTGCCCGAGAGAG 1410
DB 381 GlyTyGlnLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400
QY 1411 AGCTGGACCGTGAAGCATCTCAGAGCTGTGGCGAAGCTGAAGTGGCGCGCGCGCGCGCG 1470
DB 401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420
QY 1471 TACCCGCGCATCAAGTGGCGCGCGCTGTGCAAGCTGTGCGCGCGCGCGCGCGCGCGCG 1530
DB 421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440
QY 1531 GACATCGTGGCGCTGACCGAGGCGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTG 1590
DB 441 AspIleValProLeuThrGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeu 460
QY 1591 CGCGAGCCGTGCGAGGCGGTGTACTACGACCCCGAGAGGAGCTGTGGCGCGAGATCTCCAG 1650
DB 461 LysGluProValHisGlyValTyTyAspProSerLysAspLeuIleAlaGluIleGln 480
QY 1651 AAGCAGGCGCACCGAGTGGACCTACAGATCTACAGAGAGCGCTTCAAGAACCTGTAAG 1710
DB 481 LysGlnGlyAspAspGlnTrpThrTyGlnIleTyGlnGluProPheLysAsnLeuLys 500
QY 1711 ACCGCGCAAGTACGCGAAGATCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1770
DB 501 ThrGlyLysTyAlaLysArgThrThrHisThrAsnAspValLysGlnLeuThrGlu 520
QY 1771 GCGGTGCGAAGATCGCGATCGAGAGCATCGTGATCTGGGCGAAGACCCCAAGTTCGCG 1830
DB 521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540
QY 1831 CTGCGCATCAAGAGAGACCTGGAGACCTGTGGACCGCTACTGCGAGCGCGCACCTGG 1890
DB 541 LeuProIleGlnLysGluThrTrpGluIleTrpIleTrpIleTrpIleTrpIleTrpIle 560
QY 1891 ATCCCGAGTGGGAGTTCTGTAACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1950
DB 561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTyTrpTyGlnLeuGlu 580
QY 1951 AAGAGGCCATCATCGCGCGCGAGACCTTCTACGTGAGCGCGCGCGCGCGCGCGCGCG 2010
DB 581 LysGluProIleAlaGlyAlaGluThrPheTyValAspGlyAlaAlaAsnArgGluThr 600
QY 2011 AAGATCGGCAAGCGCGCTACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2070
DB 601 LysIleGlyLysAlaGlyTyThrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620
QY 2071 GAGACCAACCAAGAGACCGAGCTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2130
DB 621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640
```



```

QY 1351 ---CCCATCGAGTGCACCCGACAAAGTGGACCGTCAGCCCATCGAGTGCCTCCGAGAAG 1407
Db |||||||
QY 520 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLys 539
Db |||||||
QY 1408 GAGAGCTGGACCGTGAACGACATCCAGAACTGGTGGCAAGCTGAACTGAGCTGGGCCCAAGCAG 1467
Db |||||||
QY 540 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpThrSerGln 559
Db |||||||
QY 1468 ATCTACCCCGGATCAAGTGGCCAGCTGCTCAAGCTGCTGGCGGGCCGCAAGCCCTG 1527
Db |||||||
QY 560 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 579
Db |||||||
QY 1528 ACCGACATCGTCCCTGACCGAGAGGCGGAGCTGGAGCTGGCGAGAACCGCGAGATC 1587
Db |||||||
QY 580 ThrAspIleValProLeuThrGluGlnAlaGluLeuGluLeuAlaGluAsnArgGluLe 599
Db |||||||
QY 1588 CTGCGCGAGCCCGTGCACGGCGTGTACTACGACCCCGAGCAAGCACTGCTGGTGGCCGAGATC 1647
Db |||||||
QY 600 LeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluLe 619
Db |||||||
QY 1648 CAGAGCAGGCGCACGACAGTGGACCTACAGATCTACGAGAGCCCTTCAAGAACCTG 1707
Db |||||||
QY 620 GlnLysGlnGlyAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 639
Db |||||||
QY 1708 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGACGCTGAAGCAGCTGACC 1767
Db |||||||
QY 640 LysThrGlyLysTyrAlaLysArgThrThrHisThrAsnAspValLysGlnLeuThr 659
Db |||||||
QY 1768 GAGGCGGTGCAGAGATCGCCATGGAGAGCACTCGTATCTGGGGCAAGACCCCAAGTTC 1827
Db |||||||
QY 660 GluAlaValGlnLysIleSerLeuGluSerIleValThrTrpGlyLysThrProLysPhe 679
Db |||||||
QY 1828 CCCTGCGCCATCCAGAGAGACCTGGAGACCTGGTGACCGACTACTGCGAGGCCACC 1887
Db |||||||
QY 680 ArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThr 699
Db |||||||
QY 1888 TGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGTACCAGCTG 1947
Db |||||||
QY 700 TrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 719
Db |||||||
QY 1948 GAGAGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCAG 2007
Db |||||||
QY 720 GluLysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 739
Db |||||||
QY 2008 ACCAGATCGCAAGCGCGGTACTGTACCGACCGCGGCGCGCAGAGATCGTGACCTG 2067
Db |||||||
QY 740 ThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeu 759
Db |||||||
QY 2068 ACCGAGACCCACCAACGAGACCGAGCTGCAGGCGCATCCAGCTCGCCCTGCAGGACAGC 2127
Db |||||||
QY 760 SerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 779
Db |||||||
QY 2128 GGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCGCTGGGCATCATCCAGGCCAG 2187
Db |||||||
QY 780 GluSerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 799
Db |||||||
QY 2188 CCGCAGCAGAGCGAGCGAGCTGGTGAACGAGATCATCGACGACTGATCAAGAGGAG 2247
Db |||||||
QY 800 ProAspArgSerGluSerGluLeuValAsnGlnIleIleGlnGlnLeuIleLysLysGlu 819
Db |||||||
QY 2248 AAGGTGTACTCGTGGTGGTCCCGCCACAGGCGCATCGCGGCGCAACGAGCAGATCGAC 2307
Db |||||||
QY 820 ArgAlaTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysAspGlnValAsp 839
Db |||||||
QY 2308 AAGCTGGTGAAGCAAGGATCCCGCAAGGTGCTG 2340
Db |||||||
QY 840 LysLeuValSerSerGlyIleArgLysValLeu 850

```

RESULT 3  
 US-10-283-847-17  
 ; Sequence 17, Application US/10283847  
 ; Publication No. US20030162720A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: Nicolette, Charles A.
; APPLICANT: Walker, Bruce
; TITLE OF INVENTION: THERAPEUTIC ANTI-HIV (IV9) COMPOUNDS
; FILE REFERENCE: GZ 2112.00
; CURRENT APPLICATION NUMBER: US/10/283,847
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 60/345,116
; PRIOR FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus
; US-10-283-847-17

```

Alignment Scores:  
 Pred. No.: 4,06e-160 Length: 1003  
 Score: 3454.00 Matches: 651  
 Percent Similarity: 95.01% Conservative: 34  
 Best Local Similarity: 90.29% Mismatches: 28  
 Query Match: 75.83% Indels: 8  
 DB: 14 Gaps: 4

US-09-610-313B-32 (1-2457) x US-10-283-847-17 (1-1003)

```

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGGCAAGGCCCGGAGTTCCCGACGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCAACAGACAGCCACCGCGAGCTGCAGGTG-----CGCGGGGACAAACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCAGGCGCGCGCGAGCGCGCAGCGCCAGCCACCTG-----AACTTCCCCCAGATCACC 387
Db 41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
QY 398 CTGTGGCAGCGCCCTGTGTGAGATCAAGTGGCGGCGCAGATCAAGAGGCGCCCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGAGCACACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGGAAGCCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluMetSerLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGGCGCATCGCGGCTTTCATCAAGGTGCGCCAGTACGACCCAGATCTGTATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTCGGCGAAGAGCGCATCGGACCGTGTGATCGGCGCCCGCCCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGCTGACCTGAACTCCCATCAGCCCGCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGGCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCGCCCAAGGTGAAGCAGTGGGCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGCCCTGACCGCATCTCGGAGGAGATCGAGAGAGAGGCGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACAAGATCGGCGCGCGAGAACCCCTACAAACACCCCGCTTCCGCGCTCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240

```

Qy 928 GACTTCTGGAGGTGCTAGCTGGGATCCCTCCACCCCGCCGCTGAGAGAAAGAGC 987  
 Db 241 AspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLysLysSer 260  
 Qy 988 GTGACCGTGTGGACGTGGGCGACCTTACTTCTAGCGTCCCTGGAGGACTTCGC 1047  
 Db 261 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArg 280  
 Qy 1048 AAGTACACCGCTTCCACCATCCCGAGCATCAACACGAGACCCCGGCGATCCGCTACGAG 1107  
 Db 281 LysTrpThrAlaPheThrIleProSerIleAsnGlnThrProGlyIleArgTrpGln 300  
 Qy 1108 TACAACTGTCTGCTCCCGAGGCTGGAAGCGCAGCCCGACATCTTCCAGAGCATGACC 1167  
 Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 Qy 1168 AAGATCTCTGGAGCCTTCCGCGCCGCGACCCCGAGATCGTGATCTACCGAG-----GCC 1221  
 Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTrpGlnTyrMetAsp 340  
 Qy 1222 CCCCTGTACGTGGGCGACCTCGAGATCGCGCAGCACCGCGCAAGATCGAGGAGCTG 1281  
 Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 Qy 1282 CGCAAGCACCTGTCTGCTGGGGCTTCCACCAACCCCGACAGAGACCAAGAGGAGCC 1341  
 Db 361 ArgGlnHisLeuLeuAspGlyTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380  
 Qy 1342 CCCTTCTCTG-----CCCATCGAGCTGACCCCGACAGTGGAGCGTGGAGCCCATCGAG 1395  
 Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
 Qy 1396 CTGCGCCGAGAGGAGAGCTGACCGCTGAACACATCCAGAGCTGGTGGGCACTGAAC 1455  
 Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
 Qy 1456 TGGCGCCAGCAGATCTACCCCGCATCAAGTGGCGCCAGCTGTCAAGCTGCTCGCGGC 1515  
 Db 421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
 Qy 1516 GCCAAGCCCTGACCGCATGCTGCTCCCTGACCGAGAGGCGGAGCTGGAGTGGCGAG 1575  
 Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460  
 Qy 1576 AACCGCAGATCTCTGCGGAGCCGCTGACGCGCTGTACTACGACCCCGACGAGCCTG 1635  
 Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAspProSerLysAspLeu 480  
 Qy 1636 GTGCGCGAGATCCAGAGCAGGCGCCAGCACCGAGTGGACCTACCGATCTACCGAGGCC 1695  
 Db 481 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500  
 Qy 1696 TTCGAAGACCTGAAGACCGCAAGTACGCCAAGATGCGCACCGCCACACCAACGAGCTG 1755  
 Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
 Qy 1756 AAGCAGCTGACCGAGGCGGTGACAGATCCCATGGAGAGCATCGTGATCTGGGGCAG 1815  
 Db 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540  
 Qy 1816 ACCCGCAAGTCTCGCTGCGCCATCCAGAGGAGACCTGGGAGACTGTGTGACCGACTAC 1875  
 Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560  
 Qy 1876 TGGCAGGCGCAGCTGGATCCCGGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1935  
 Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
 Qy 1936 TGGTACCGAGCTGGAGAGAGGAGCCATCATCGCGCGCGAGACTTCTAGTGGAGCGGCGC 1995  
 Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600

Qy 1996 GCCAACCGCAGAGATCGGCAAGCGCGGTACGTGACCGACCGCGCGCGAGAG 2055  
 Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620  
 Qy 2056 ATCTGTAGCTGACCGAGACCCCAACAGAGACCGAGCTGCGAGCCATCCAGCTGGCC 2115  
 Db 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640  
 Qy 2116 CTGACGACGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCTCGGCATC 2175  
 Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
 Qy 2176 ATCCAGCGCCAGCCGACAGAGCGAGCGAGCTGTGTGAACACGATCATCGAGCGTG 2235  
 Db 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680  
 Qy 2236 ATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCCAAGGGCATCGCGGCAAC 2295  
 Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
 Qy 2296 GAGCAGATCACAGCTGTGTGAGCAAGGCGATCCGCAAGGTGTCTTCTCGAGCGCATC 2355  
 Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
 Qy 2356 GAT 2358  
 Db 721 Asp 721

RESULT 4  
 US-10-634-165-9  
 ; Sequence 9, Application US/10634165  
 ; Publication No. US2005009581A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chang, Nancy T  
 ; APPLICANT: Gallo, Robert C.  
 ; APPLICANT: Wong-Staal, Flossie  
 ; TITLE OF INVENTION: DETECTION OF HIV-1 DNA  
 ; FILE REFERENCE: 223695  
 ; CURRENT APPLICATION NUMBER: US/10/634,165  
 ; CURRENT FILING DATE: 2003-08-05  
 ; PRIOR APPLICATION NUMBER: 08/463,028  
 ; PRIOR FILING DATE: 1995-06-05  
 ; PRIOR APPLICATION NUMBER: 06/693,866  
 ; PRIOR FILING DATE: 1985-01-23  
 ; PRIOR APPLICATION NUMBER: 06/659,339  
 ; PRIOR FILING DATE: 1984-10-10  
 ; PRIOR APPLICATION NUMBER: 06/643,306  
 ; PRIOR FILING DATE: 1984-08-22  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 9  
 ; LENGTH: 1015  
 ; TYPE: PRT  
 ; ORGANISM: T cell leukemia-lymphoma virus (HTLV)  
 ; FEATURE:  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (1)-(1015)  
 ; OTHER INFORMATION: /note= "pol protein of HTLV-III"  
 US-10-634-165-9

Alignment Scores:  
 Pred. No.: 4,78e-159 Length: 1015  
 Score: 3432.00 Matches: 650  
 Percent Similarity: 93.18% Conservative: 33  
 Best Local Similarity: 88.68% Mismatches: 30  
 Query Match: 75.35% Indels: 20  
 DB: 17 Gaps: 5

US-09-610-313B-32 (1-2457) x US-10-634-165-9 (1-1015)

Qy 220 TTCTTCGCGAGACCTGGCTTCCCGCAGGCGCGCGAGTTC----- 267  
 Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

QY 268 -----CCAGCGAGCAACCGCGCAACAGCCCGCCAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
QY 304 CGCGAGCTCAGGTG-----CGCGCGCAACCCCGCAGCGAGCGCGCGCCAGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
QY 358 CAGGGCACCTCG-----NACTTCCCGCAGATCACCTGTGTGCAGGGCCCTGTGTGAC 411  
Db 61 GlnGlyThrValSerPheAsnProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
QY 412 ATCAAGGTGGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGCAGCACCGGTG 471  
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
QY 472 CTGGAGGAGATGAGCTGCCCGCAAGTGGAAAGCCCAAGATGATCGGGCGCATCGCGCGC 531  
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120  
QY 532 TTCATCAAGTGGCGCGCAGTACACCGATCCTGTATCGAGATCTGGCGCAAGAGCCATC 591  
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
QY 592 GGCACGCTGCTGATCGGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651  
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
QY 652 CTGGGTGCACCTCAACTTCGCCCATCAGCCCATCGAGACCGTGCCTGAACTGAAG 711  
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CGCGGATGGAGCGGCGCCCAAGTGAAGCAAGTGGCCCTGACCGAGGAGAAGATCAAGGCC 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCGCCATCTGGAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTACAACACCCCGCTGTCGCCATCAAGAAGAAGGACAGACCAAGTGGCGCAAG 891  
Db 221 AsnProTyAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCGGGAGCTGAACAGCGCAACCGAGACTTCTGGAGAGTGCAGCTGGGC 951  
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCCACC CGCGCTGAAGAGAGAGAGCGTGACCGTGTGACGTGGCGGCAC 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTCAGCGTGCCTGCACAGGACTTCCGCAAGTACACCGCTTACCATCCCC 1071  
Db 281 AlaTyPheSerValProLeuAspGluAspPheArgLysTyThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACAACAGACCCCGCATCCGCTACCAAGTACAAGTCTGCCCGCAGGGCTGG 1131  
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyGlnTyAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGGCAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGGCCCTTCGCGCC 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCCGAGATGCTGATCTACAG-----GCCCGCTGTAGTGGCGAGCACCTG 1245  
Db 341 GlnAsnProAspIleValIleTyGlnTyMetAspAspLeuTyValGlySerAspLeu 360  
QY 1246 GAGATCGGCAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGCG 1305  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380

QY 1306 TTCACACCCCGCAGCAAGAAGCACAGAGAGCCCGCTTCTCTG-----CCCATCGAG 1359  
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyTrpGlu 400  
QY 1360 CTGACACCCCGCAAAAGTGGACCGTCCAGCCATCGAGTGTCCCGAGAAAGGAGAGCTGGACC 1419  
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1420 GTGAACGATCCAGAGCTGGTGGGCAAGCTGAAGTGAAGTGGGCGAGCCAGATCTACCCCGC 1479  
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyTrpGly 440  
QY 1480 ATCAAGGTGGCCCGAGCTGTCAAGCTGTCCGCGCGCGCAAGGCCCTCACGACATCGTG 1539  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1540 CCGCTGACCGAGGAGCGAGCTGGAGTGGCCGAGAACCGCGAGATCCTGGCGAGGCC 1599  
Db 461 ProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1600 GTGCACGGCTGTACTAGCACCCAGCAAGACCTGGTGGCGGAGATCCAGAACGAGGCC 1659  
Db 481 ValHisGlyValTyTrpAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1660 CACGACCAAGTGCACCTACAGATCTACCAAGAGCCCTTCAAGAACCTCAAGACCGGCAAG 1719  
Db 501 GlnGlyGlnTrpThrTyThrGlnIleTyThrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1720 TACCGCAAGATGCGCACCGCCACACCAACGACGTGAAGACGCTGACCGAGCGCGTGCAG 1779  
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1780 AAGATCGCCATGGAGAGATCTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATC 1839  
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1840 CAGAGGAGACCTGGGAGACCTGTGGACCTGACCTACTGCGAGGCGCACCTGATCCCGCAG 1899  
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyTrpGlnAlaThrTrpIleProGlu 580  
QY 1900 TGGAGTTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTACCAAGTGGAGAGGAGGCC 1959  
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyThrGlnLeuGluLysGluPro 600  
QY 1960 ATCATCGCGCGCAGACCTTCTACGTGAGCGCGCGCGCAACCGCGAGACCAAGATCGGC 2019  
Db 601 IleValGlyAlaGluThrPheTyValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2020 AAGCGCGCTACGTGACCGACCGCGCGCGCAGAGATCGTGAGCTGACCGAGACCAACC 2079  
Db 621 LysAlaGlyTyTrpValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2080 AACCAGAGACCGAGCTGCAGGCCATCCAGCTGGCGCTGACGAGGACAGCGGCGAGGAGTG 2139  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2140 AACATCGTGAACGACGACGACGATCGCTGGGCGATCATCCAGCGCGCGCAGCCCGCAAGAGC 2199  
Db 661 AsnIleValThrAspSerGlnTyAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2200 CAGAGCGAGCTGGTGAACCAACGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTG 2259  
Db 681 GluSerGluLeuValAsnGlnIleIleGlnLeuIleLysLysGluLysValTyTrpLeu 700  
QY 2260 AGCTGGGTGGCCCGCCCAAGGGCATCGCGCGCAACGAGACGATCGACAAGCTGTGTGAGC 2319  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyLysGlnGluValAspLysLeuValSer 720  
QY 2320 AAGGCAATCCGCAAGGTGCTGTTCTGAGCGCGCATCGAT 2358  
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 5

US-10-093-953A-26

; Sequence 26, Application US/10093953A

; Publication No. US20040105871A1

; GENERAL INFORMATION:

; APPLICANT: Robinson, Harriet L.

; APPLICANT: Smith, James M.

; APPLICANT: Hua, Jian

; APPLICANT: Moss, Bernard

; APPLICANT: Amara, Rama

; APPLICANT: Wyatt, Linda

; APPLICANT: Earl, Patricia

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING

; FILE OF INVENTION: AN IMMUNE RESPONSE

; FILE REFERENCE: 12804-005002

; CURRENT APPLICATION NUMBER: US/10/093.953A

; CURRENT FILING DATE: 2002-03-08

; PRIOR FILING DATE: US 60/186,364

; PRIOR FILING DATE: 2000-03-02

; PRIOR FILING DATE: US 60/251,083

; PRIOR FILING DATE: 2000-12-01

; PRIOR APPLICATION NUMBER: US 09/798,675

; PRIOR APPLICATION NUMBER: PCT/US01/06795

; PRIOR FILING DATE: 2001-03-02

; PRIOR FILING DATE: US 60/324,845

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: US 60/325,004

; PRIOR FILING DATE: 2001-09-26

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 26

; LENGTH: 739

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: protein encoded by construct of vaccine vector

; OTHER INFORMATION: pGA2 and insert JS2 expressing clade HIV-1 VL

US-10-093-953A-26

Alignment Scores:

Pred. No.:	6.15e-158	Length:	739
Score:	3409.00	Matches:	647
Percent Similarity:	92.91%	Conservative:	34
Best Local Similarity:	88.27%	Mismatches:	32
Query Match:	74.84%	Indels:	20
DB:	16	Gaps:	5

US-09-610-313B-32 (1-2457) x US-10-093-953A-26 (1-739)

Qy	220	TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGCAGCGAGCAG	279
Db	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
Qy	280	AACGCG-----GGCAACAGCCCGCCAGC	303
Db	21	ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg	40
Qy	304	CGCGAGCTGCAGTG-----CGCGGCAACCCCGCAGCGAGCGCGCGCGAGCGC	357
Db	41	ArgGluLeuGlnValTrpGlyArgAspAsnSerProSerGluAlaGlyAlaAspArg	60
Qy	358	CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGGTGAGC	411
Db	61	GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr	80
Qy	412	ATCAAGTGGCGCGCCAGATCAAGAGCGCTGCTGGACACCGCGCCGCGAGCACCGTG	471
Db	81	IleLysIleGlyGlnLeuLysGluAlaLeuAspThrGlyAlaAspThrVal	100
Qy	472	CTGGAGAGATGACCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC	531
Db	101	LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGly	120

Qy	532	TTCAATCAAGTGGCGCCAGTAGCAGACCATCTGATCGAGATCTCGCGCAAGAGCGCCATC	591
Db	121	PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle	140
Qy	592	GGCAGCGTGTGATCGCGCCCGCCCGCTGAACATCATCGCGCGCACATCTGACCCAG	651
Db	141	GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln	160
Qy	652	CTGGGCTGCACCCCTGAACCTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAG	711
Db	161	IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys	180
Qy	712	CCGCGCATGAGCGCCCGCCAGGTGAACAGTGGCCCTGACCGAGAGAGATCAAGGCC	771
Db	181	ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla	200
Qy	772	CTGACCGCCCATCTGCGAGGAGATCGAGAGAGGCGCAAGATCACCAAGATCGGCCCGAG	831
Db	201	LeuValGluIleCysThrGluMetGluLysIleSerLysIleGlyProGlu	220
Qy	832	AAGCCCTTACCAACACCCCGCTGTTCGCCCATCAAGAGAGAGACACCAAGTGGCGCAAG	891
Db	221	AsnProTyraAsnThrProValPheAlaIleLysLysLysSerThrLysTrpArgLys	240
Qy	892	CTGGTGGACTTCCCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGCG	951
Db	241	LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly	260
Qy	952	ATCCCCCACCCCGCGCTTGAAGAGAGAGAGCGTGCCTGCTGGAGCTGGCGGAC	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp	280
Qy	1012	GCCTACTTTCAGCGTGCCTGGAGGAGCTTCGCGAAGTACACCGCTTCCACCATCCCC	1071
Db	281	AlaTyraPheSerValProLeuAspGluAspPheArgLysTrpThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACAAACGAGACCCCGCGCATCCGTTACAGTACAACTGCTGCCCGAGGCTGG	1131
Db	301	SerIleAsnAsnGluThrProGlyIleArgTyraGlnTyraAsnValLeuProGlnGlyTrp	320
Qy	1132	AAGGCGAGCCCGCAGCATCTTCCAGAGGAGCATGACCAAGATCTGGAGCGCTTCCGCGC	1191
Db	321	LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys	340
Qy	1192	CGCAACCCCGAGATCGTGATCTACCAG-----GCCCGCTGTACGTGGCGAGGAGCTG	1245
Db	341	GlnAsnProAspIleValIleTyraGlnTyraMetAsnAspLeuTyraValGlySerAspLeu	360
Qy	1246	GAGATCGCGCAGCACCGCGCCAAAGATCGAGAGCTGCGCAAGACCTGCTGCTGGCGGC	1305
Db	361	GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly	380
Qy	1306	TTCAACACCCCGCAGAGAGCAGCAGAGAGCGCCCGCTTCCCTG-----CCCATCGAG	1359
Db	381	LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyraGlu	400
Qy	1360	CTGCACCCCGCAGCAAGTGGACCGTGCAGCTGCAGCTGCCGAGAGAGAGAGTGGAGC	1419
Db	401	LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr	420
Qy	1420	GTGAACGACATCCAGAGCTGGTGGGCAAGCTGAACCTGGCGCCAGCCAGATCTACCCCGC	1479
Db	421	ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyraProGly	440
Qy	1480	ATCAAGTGGCGCAGCTGTGCAAGCTGCTCGCGCGCCCAAGCGCTGACCGCATCGTG	1539
Db	441	IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle	460
Qy	1540	CCCTGACCGAGGAGCGCGAGCTGGAGCTGCCGAGAACCGCGAGATCTCTGCGCGAGCCC	1599
Db	461	ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro	480
Qy	1600	GTGCACGGCGTGTAACGACCCCGCAGAGACCTGTGTGGCGCAGATCCAGAGAGCGGC	1659



```
Db 481 ValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1660 CACGACCACTGAGCTACAGATCTACAGAGAGCCCTTCAAGAACTGAAGACCGGCAAG 1719
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1720 TACGCCAAGATGCGCACCGCCACACCAACGACGTAAGCAGCTCACCGAGGCGGTGAG 1779
Db 521 TyrAlaArgMetCysGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
QY 1780 AAGATCGCCATGAGAGATCTGTGCTGGGCAAGACCCCAAGTTCGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1840 CAGAGGAGACTGGAGACCTGTGGACCGACTACTGGACGACTACTGGACGACTGATCCCGAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpThrTrpThrGluThrTrpGlnAlaThrTrpIleProGlu 580
QY 1900 TGGGAGTTCTGTAACACACCCCTGCTGAGCTGTGTACCATCTGGAGAGGAGGCC 1959
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1960 ATCATCGGCGCGACACCTTCTACCTGACGCGCGCGCCCAACCGGAGACCAAGATCGCG 2019
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCGCGCTACTGACACGACCGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCCACC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACCAAGACCGACGCTGAGCCATCCAGCTGGCGCTGCGAGGACGCGGACGAGCGAGTG 2139
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2140 AACATCGTACCGACGACGACGACGCGCTGCGCATCATCCAGCCCGACGCGGCAAGAGC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGCGGAGCTGGTGAACAGATCATCGAGAGCTGATCAAGAGAGAGGTGTACTGT 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTGCGCGCCCAAGGGCATCGCGCGCAAGCAGCAGATCGACAGAGTGTGAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2320 AAGGGCATCCGCAAGTGTGTCTTCTGACGCGCATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

## RESULT 6

```
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: ROBINSON, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE REFERENCE: 12804-005002
; CURRENT APPLICATION NUMBER: US/10/093,953A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein encoded by construct of vaccine vector
; OTHER INFORMATION: pCAL and vaccine insert expressing clade B HIV-1
; OTHER INFORMATION: Gag-Pol
US-10-093-953A-31
Alignment Scores: 6.15e-158 Length: 739
Score: 3409.00 Matches: 647
Percent Similarity: 92.91% Conservative: 34
Best Local Similarity: 88.27% Mismatches: 32
Query Match: 74.84% Indels: 20
DB: 16 Gaps: 5
US-09-610-313B-32 (1-2457) x US-10-093-953A-31 (1-739)
QY 220 TTCTTCCCGAGGACCTGGCTTCCCGGCAAGGCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACGCG-----GCCAACGCCCCACCCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAAGTG-----CGCGCGACAAACCCCGCAGCGCGCGCGCGCGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGTGCGAGCGCCCTGTGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGTGGCGCGCGCAGATCAAGAGCGCTGTGTGACACCGCGCGCGCGCGCGCGCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100
QY 472 CTGGAGGAGATGAGCTCGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGly 120
QY 532 TTCATCAAGTGGCGCGCAGTACGACACAGATCTCTGATCGAGATCTCGCGCAAGAGCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCTGAACCTCCCATCAGCCCATCAGCGCGCGCGCGCGCGCGCGCGCGCG 711
Db 161 IleGlyCysThrLeuAsnPheProLysSerProIleGluThrValProValLysLeuLys 180
QY 712 CCGGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCGCGCTCTGGAGGAGATGAGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
```

```
QY 832 AACCCCTACACACCCCGCTGTTCCCATCAAGAAAGAGCAGCACCACCAAGTGGCGCAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGTGTGACTTCGCGAGCTGAACAAGCAGCACCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCCACCCCGCGCTGAAGAAGAAGAGCGTGCACCGTGTGAGCGTGGCGCAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTCAGCTGCGCTGGAGAGGACTTCGCGAAGTACACCGGCTTCACCATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACAGCAGACCCCGGATCCGCTACCAAGTACACAGTGCCTGCCCGCGCTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGCGAGCCCGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
QY 1192 CGCAACCCCGAGATCGTGATCTACAG-----GCCCGCTGTAGTGGGAGGAGCCTG 1245
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360
QY 1246 GAGATCGCGCAGCAGCCGCGCAGATCGAGGAGCTGCGCAGCAGCTGCGCTGGCGGC 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuA-gtrpGly 380
QY 1306 TTCACACCCCGCAAGAACAGCAGCAGAGGAGCCCGCTTCTCTG-----CCCATCGAG 1359
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1360 CTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGAGCTGAC 1419
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1420 GTGACACACATCCAGAGCTGTGGCAGCTGAAGTGGCGCCAGCAGATCAACCCCGC 1479
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440
QY 1480 ATCAAGTGGCGCAGCTGTGAAGCTGCTGCGCGCGCAGGCGCTGACCGACCATCGT 1539
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1540 CCCGTGACCGAGGAGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGCGAGCC 1599
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1600 GTGACGCGGTGTACTAGCACCAGCAGACCTGTGTGGCGGAGATCCAGAGCAGGCG 1659
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1660 CACGACCATGTGCCTTACCATCTTACAGAGCGCTTCAAGAACCTTGAAGACCGCAAG 1719
Db 501 GlnGlyGlnThrThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1720 TACGCCAAGATGCGACCGCGCCACACCAAGCTGAAGCAGCTGACCGAGCGCGCTGAG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
QY 1780 AAGATCGCATGGAGAGCATGTGATCTGGGGGAAGACCCCAAGTTCGCGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1840 CAGAAGAGACCTGGGAGACCTGTGGACCGACCTACTGGGAGCGCCACCTGATCCCCGAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpTrpTrpGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1900 TGGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGTGACCATCGTGGAGAGGAGGCC 1959
```

```
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1960 ATCATCGCGCCGAGACCTTCTACGTGGAGCGGCGCCCAACCGCGAGACCAAGATCGGC 2019
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCGCGCTACGTGACCGACCGCGCGCGCGAGAGATCGTGGCTGACCGAGCAGCACC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACGAGAGACCGAGCTGCGAGGCGCATCCAGCTGCGCGCGAGCAGCAGCGCGAGGTG 2139
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2140 AACATCGTGAACCGACCGCAGTACCGCTGGCATCATCCAGGCCCGACCGCAAGAGC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGAGCAGCTGTGTGAACAGCATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTGT 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTGCGCGCCACAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGTGAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGlnGlnValAspLysLeuValSer 720
QY 2320 AAGGCGATCCGCAAGGTGTCTCTCGCGCGCATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7
US-10-325-468-23
; Sequence 23, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; FILE REFERENCE: MODELS OF HIV-1 PATHOGENESIS
; CURRENT APPLICATION NUMBER: US/10/325,468
; PRIOR FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.10 protein Pol
US-10-325-468-23

Alignment Scores:
Pred. No.: 7,38e-157 Length: 1003
Score: 3387.00 Matches: 634
Percent Similarity: 94.17% Conservative: 45
Best Local Similarity: 87.93% Mismatches: 34
Query Match: 74.36% Indels: 8
DB: 16 Gaps: 4

US-09-610-313B-32 (1-2457) x US-10-325-468-23 (1-1003)
QY 220 TTCTTCGCGAGGACCTTGGCTTCCCGAGGCAAGCGCCCGAGTTCCTCCGAGGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
QY 280 AACCGCGCCACACCGCCCGAGCTGAGGTG-----CGCGGCGACACACCC 333
Db 21 ThrArgAlaAsnSerProIleArgGluGlnValTrpArgA-gspAsnAsnSer 40
```



APPLICANT: Soong, Nay Wei  
APPLICANT: Pekrun, Katja  
APPLICANT: Shibata, Riri  
TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL  
FILE REFERENCE: 0166.210US  
CURRENT APPLICATION NUMBER: US/10/325,468  
CURRENT FILING DATE: 2002-12-19  
PRIOR APPLICATION NUMBER: US 60/343,524  
PRIOR FILING DATE: 2001-12-21  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 35

LENGTH: 1003

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol

US-10-325-468-35

#### Alignment Scores:

Pred. No.:	9,23e-157	Length:	1003
Score:	3385.00	Matches:	635
Percent Similarity:	94.17%	Conservative:	44
Best Local Similarity:	88.07%	Mismatches:	34
Query Match:	74.31%	Indels:	8
DB:	16	Gaps:	5

US-09-610-313B-32 (1-2457) x US-10-325-468-35 (1-1003)

Qy	220	TTCTTCGCGAGGACCTGGCTTCCCTCCAGGGCAAGCCCGCGAGTTCCTCCAGCGAGCAG	279
Db	1	PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln	20
Qy	280	AACCGCGCCACACGCCCTCCAGCGCGAGCTGAGTG---CGCGGGAC---AACCCC	333
Db	21	ThrArgAlaAsnSerProIleArgGluArgGlnValTrpArgGlyAspAsnSer	40
Qy	334	CGACGCGAGGCGCGCGCGAGCGCGAGCGCACCTCGT-----AACTTCCCGAGATCAC	387
Db	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr	60
Qy	388	CTGTGGCAGCGCCCTCGGTGATCAAGTGGCGGCCGAGATCAAGGAGCGCCCTGCTG	447
Db	61	LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeu	80
Qy	448	GACACCGCGCGAGCACCGCTCTGAGAGATGAGCTGCCCGGCAAGTGAAGCCC	507
Db	81	AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro	100
Qy	508	AAGATGATCGCGGCATCGCGGCTTCATCAAGTGGCGGAGTACGACCGAGATCCTGATC	567
Db	101	LysMetIleGlyIleGlyGlyPheIleLysValArgGlnIleAspGlnIleProIle	120
Qy	568	GAGATCTCGCGCAAGAGCGCATCGGACCGTGTGATCGCGCCCGCCCGCGTGAACATC	627
Db	121	AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle	140
Qy	628	ATCGCGCGCAACATGCTGACCCAGCTGGTGTGACCTGAACTTCCCATCAGCCCCATC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
Qy	688	GAGACCTGCGCTGAGCTGAAGCCCGGATGACCGCCCGCCCGAGGTGAAGCTGGCCC	747
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
Qy	748	CTGACCGGAGGAGATCAAGCGCCCTGACCGCCATCTCGGAGGAGATGAGAGAGGGC	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
Qy	808	AAGATCAACCAAGATCGGCGCGGAGAACCCCTTACACACCCCGGTGTTCCCATCAAGAAG	867
Db	201	LysIleSerLysIleGlyProGluAsnProIleAsnThrProValPheAlaIleLysLys	220

Qy	868	AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCCACCCAG	927
Db	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln	240
Qy	928	GACTTCTGGGAGGTGAGCTGGGCATCCCCACCCCGCGCGCTGAAGAAGAAAGAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
Qy	988	GTGACCGTGTGGACGTGGCGCGACCTACTTACGCTGCCCTGGAGCAGGAGCTTCGC	1047
Db	261	ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspLysAspPheArg	280
Qy	1048	AAGTACACCGCCTTACCATCCCGAGCATCAACAACGAGACCCCGCGCATCCCGTACCAG	1107
Db	281	LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln	300
Qy	1108	TACAACTGTGCTGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC	1167
Db	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320
Qy	1168	AAGATCTGTGAGCCCTTCCGCGCGCAACCCCGAGATCGTGATCTACCAG-----GCC	1221
Db	321	LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleIleIleIleIleIleIle	340
Qy	1222	CCCTGTGTACGTGGCGAGCACCTGGAGATCGCGCAGCACCGCGCAAGATCGAGGAGCTG	1281
Db	341	AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
Qy	1282	CGCAAGCCTGTGTGCGCTGGGCTTCCACACCCCGCAAGAACGACCAAGAGAGGCC	1341
Db	361	ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
Qy	1342	CCCTTCTCTG-----CCCATCGAGCTGCACCCGCAAGTGGACCGCGAGCCCATCGAG	1395
Db	381	ProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal	400
Qy	1396	CTGCCCGAGAACGAGCTGCAGCCGTGAACGACATCCAGAACGTGGTGGGCAAGCTGAAC	1455
Db	401	LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
Qy	1456	TGGGCCAGCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGTGTCTGCGCGC	1515
Db	421	TrpAlaSerGlnIleTrpAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly	440
Qy	1516	GCCAAAGCCCTGACCGACATCGTCCCTCACCGAGAGCGCGAGCTGGAGCTGGCGCAG	1575
Db	441	ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu	460
Qy	1576	AACCGCGAGATCTGCGCGAGCCCGTGCACGGCTGTACTACGACCCCGAGCAAGGACCTG	1635
Db	461	AsnArgGluLeuLeuLysGluProValHisGlyValTrpTrpAspProSerLysAspLeu	480
Qy	1636	GTGCGCGAGATCCAGAACGAGCGCCACGACGAGTGGACCTTACCAATCTTACCGAGAGCC	1695
Db	481	IleValGluIleGlnLysGlnGlyGlnTrpThrTrpThrTrpGlnIlePheGlnGluPro	500
Qy	1696	TTCAAGAACCTGAAGCCGCAAGTAGCCCAAGATGGCGACCCCGCCACCAAGAGAGTG	1755
Db	501	PheLysAsnLeuLysThrGlyLysTrpAlaLysThrArgGlyAlaHisThrAsnAspVal	520
Qy	1756	AAGCAGCTGACCGAGCGCTGCAGAGATGCCATGGAGAGCATCGTGATCTGGGGCAAG	1815
Db	521	LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys	540
Qy	1816	ACCCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGAGCCTGGTGGACCGACTAC	1875
Db	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpGluTyr	560
Qy	1876	TGGCAGGCCACCTGGATCCCGAGTGGGAGTTGTTGTAACACCCCGCTGGTGGTGAAGCTG	1935
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu	580

QY 1936 TGTACACAGCTGGAGAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGACGCGGCC 1995  
 Db 581 TTPTYRGLNLEUGLULYSGLUPROILLEVALGLYALAGLUTHRPHELYRVALASPGLYALA 600  
 QY 1996 GCGACCCGCGAGACCAAGATCGGCAAGCGCGGCTACGTACCGACCGGGCCGCGAGAG 2055  
 Db 601 ALAASNARGGLUTHRILYSLEUGLYLYSALAGLYTYRVALTHRSERARGGLYARGGLNLYS 620  
 QY 2056 ATCGTGAGCTCAGCGAGACCAACACAGAGAGCCGAGCTCAGGCCATCCAGCTGGCC 2115  
 Db 621 VALVALSERLEUTHRSPTTHRRASNGLNLYSTHRLULEUGLINALALLEHISLEUALA 640  
 QY 2116 CTCGAGGACGCGGCGAGGTGAACATCTGTACCCGACAGCGACGTACGCCCTGGGCGATC 2175  
 Db 641 LEUGINASPSERGLYLEUGLULVALAENILLEVALTHRASPSERGLNLYRALALEUGLYILE 660  
 QY 2176 ATCCAGGCCCGCGACCAAGACGAGAGCGAGCTGTGTGAACACAGATCATCGAGCGCTG 2235  
 Db 661 ILEGINALAGLNPROASPLYSERGLUSERGLULEUVALSERGLNILEILEGLUGLNLEU 680  
 QY 2236 ATCAAGAGGAGAGGTGTACTGTAGCTGGGTGCGCGCCGCCACAGGGCATCGCGCGCAAC 2295  
 Db 681 ILELYSLYSGULYSLVALTYRLEUTHRTPLLEPROALAHISLYSGLYLEGLYGLYASN 700  
 QY 2296 GAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTTCTCTGGACGCGCATC 2355  
 Db 701 GLUGINVALASPLYSLEUVALSERALAGLYILEARGVALLEUPHELEUASPGLYILE 720  
 QY 2356 GAT 2358  
 Db 721 Glu 721

RESULT 9  
 US-10-325-468-9  
 ; Sequence 9, Application US/10325468  
 ; Publication No. US20040101823A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Soong, Nay Wei  
 ; APPLICANT: Pekrun, Katja  
 ; APPLICANT: Shibata, Riri  
 ; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL  
 ; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS  
 ; FILE REFERENCE: 0166.210US  
 ; CURRENT APPLICATION NUMBER: US/10/325,468  
 ; CURRENT FILING DATE: 2002-12-19  
 ; PRIOR APPLICATION NUMBER: US 60/343,524  
 ; PRIOR FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 56  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9  
 ; LENGTH: 1003  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;  
 ; OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;  
 ; OTHER INFORMATION: clone P10.26 protein Pol  
 US-10-325-468-9

Alignment Scores:  
 Pred. No.: 1,03e-156 Length: 1003  
 Score: 3384.00 Matches: 634  
 Percent Similarity: 94.04% Conservatives: 44  
 Best Local Similarity: 87.93% Mismatches: 35  
 Query Match: 74.29% Indels: 8  
 DB: 16 Gaps: 4  
 US-09-610-313b-32 (1-2457) x US-10-325-468-9 (1-1003)  
 QY 220 TTCCTCCGAGAGACTCGCTGCGCTTCCCGCAGGGCAGGCGCGCGAGTTCCCGCAGGAGC 279  
 Db 1 PhePheArgGluAspLeuAlaPheProGlnGlnLYsAlaAArgLYsPheSerSerGluGln 20

QY 280 AACCGCGCAACAGCGCCACCGCGCGAGCTGCAGGTG-----CGCGGCGACAAACCCC 333  
 Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAsnAsnSer 40  
 QY 334 CGCAGCGAGCG 387  
 Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
 QY 388 CTGTGGCAGCGCCCGCTGCTGAGCATCAAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 447  
 Db 61 LeutrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysLeuAlaLeuLeu 80  
 QY 448 GACACCG 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCG 567  
 Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120  
 QY 568 GAGATCTCGCGCAAGNAGGCCCATCGGCACCGCTGCTGATCGCGCGCGCGCGCGCGCGCG 627  
 Db 121 AspileCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGCGCGCAACAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687  
 Db 141 IleglyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACCGTGCCTGTGAAGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGGAGAGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCACCAAGATCG 867  
 Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
 QY 868 AAGGACACCAAGTGGCGCAAGCTGTGTGAGCTTCCGCGAGCTGAAACAAGCGCGCGCG 927  
 Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240  
 QY 928 GACTTCTGGAGGTGCAGCTGGGCATCCCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 987  
 Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGGAGCTGGCG 1047  
 Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
 QY 1048 AAGTACACCGCTTACCATCCCGAGCATCAACAACGAGACCCCGCGCGCGCGCGCGCGCG 1107  
 Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
 QY 1108 TACAACGTGCTGCG 1167  
 Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 QY 1168 AAGATCTGTGAGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1221  
 Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340  
 QY 1222 CCCCTGTACGTGGCGAGCGACCTCGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1281  
 Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 360  
 QY 1282 CGCAAGACCTGCTGCGCTGGGGCTTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1341  
 Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCCTGT-----CCCATCGAGCTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1395

```
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
Qy 1396 CTGCCCGAAGAGGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAAC 1455
Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
Qy 1456 TGGGCCAGCAGATCTACCCCGGATCAAGTGGCGCAGCTGTGCAAGCTCTCTGGGGC 1515
Db 421 TrpAlaSerGlnIleTyrAlaGlyLysValLysGlnLeuCysLysLeuLeuArgGly 440
Qy 1516 GCCAAGGCCCTGACCGCATCGTCCCTGACCGAGGAGGCGGAGCTGGAGCTGGCCGAG 1575
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460
Qy 1576 AACCGCGAGATCTCGCGAGCCCGTGCACGCGCTGTACTACGACCCCGAGCAAGACCTG 1635
Db 461 AsnArgGlnIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
Qy 1636 GTGCCCGAGATCCAGAGCAGGCGCACGACAGTGGACCTACCATGATCTACGAGGCC 1695
Db 481 IleValGlnIleGlnLysGlnGlyGlnGlnGlnTrpThrTyrGlnIlePheGlnGluPro 500
Qy 1696 TTCAAGAACCTGAAGACCGGACGATGACCGCAAGATGCGCACCGCCACCAACGACGTG 1755
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520
Qy 1756 AAGCAGCTGACCGAGGCGCGTGCAGAAATCCCATGAGAGCATCGTATCTGGGGCAAG 1815
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaGlnSerIleValIleTrpGlyLys 540
Qy 1816 ACCCCCAAGTTCGCGTCCCATCCAGAAAGAGAGCTGGGAGACTGGTGACCGACTAC 1875
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
Qy 1876 TGGCAGGCCACTCGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTG 1935
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
Qy 1936 TGGTACCACTGAGAGAGGAGCCCATCATCGCGCCGAGACTTCTACGTGGAGCGGCGC 1995
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
Qy 1996 GCCAACCGCAGACCAAGATCGGCAAGCCCGCTACGTGACCGACCGCGGCGCGCAGAG 2055
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620
Qy 2056 ATCGTGAGCTGACGAGACCAACCAAGAGACCGAGCTGGAGGCGCATCCAGCTGGCC 2115
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
Qy 2116 CTGCAGGACAGCGGAGCGAGGTGAACATCGTACCGACGACGACGAGTACGCGCTGGCATC 2175
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Qy 2176 ATCCAGGCCCGCAGCAGAGGAGGAGCGAGCTGGTGAACACCATCATCCAGCAGCTG 2235
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
Qy 2236 ATCAAGAGGAGAGGTGTACTGAGCTGGGTGGCGCCGCGCAGAGGCGATCGGGCGAAC 2295
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
Qy 2296 GAGCAGATCGACAGCTGTGAGCAAGGGCATCCGCAAGGTGCTTCTCGAGCGGCATC 2355
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720
Qy 2356 GAT 2358
Db 721 Glu 721
RESULT 10
US-10-325-468-20
```

```
; Sequence 20, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; PRIOR FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20
```

```
Alignment Scores:
Pred. No.: 1,03e-156 Length: 1003
Score: 3384.00 Matches: 634
Percent Similarity: 94.04% Conservative: 44
Best Local Similarity: 87.93% Mismatches: 35
Query Match: 74.29% Indels: 8
DB: 16 Gaps: 4
```

US-09-610-313B-32 (1-2457) x US-10-325-468-20 (1-1003)

```
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
Qy 280 AACCGCGCAACAGCCCGCACCGCGAGCTGCAGGTG-----CGCGGCGCAACACCC 333
Db 21 ThrArgAlaAsnSerProIleArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
Qy 334 CGCAGCAGCGCGCGCGCGAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
Qy 388 CTGTGGCAGCGCCCGCTGGTGAAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
Qy 448 GACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
Qy 508 AAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
Db 101 LysMetIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 120
Qy 568 GAGATCTGCGCAAGAAGCGCATCGGACCGCTGTGATCGCGCGCGCGCGCGCGCGCGCGCG 627
Db 121 AspileCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
Qy 628 ATCGCGCGCAACATGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
Qy 688 GAGACCGTGGCGTGAAGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
Qy 748 CTGACCGGAGGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValIleCysThrGluMetGluLysGluGly 200
```





Db 21 ThrGlyAlaAenSerSerAlaSerArgLysLeu-----GlyAspGly----- 34  
Qy 340 GAGGCGCGCGCGAGCGCCAGCGCACC-----CTGAATCTTCCCCAGATCACCCCTG 390  
Db 35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53  
Qy 391 TGGCAGCGCCCTTGGTGTGAGCATCAAGGTGGGCGGCGCAGATCAAGAGGAGCGCTCTGGAC 450  
Db 54 TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 73  
Qy 451 ACCGCGCGCGAGCACCGCTGGAGGAGATGAGCTTCCCGGCAAGTGAAGCCCAAG 510  
Db 74 ThrGlyAlaAspThrValLeuGluAspIleAsnLeuProGlyLysIleThrLysProLys 93  
Qy 511 ATGATCGCGCGCATCGCGGCTTCAATCAAGGTGGCGCAGTACGACACCATCTCATCGAG 570  
Db 94 MetIleGlyIleGlyGlyPheIleLysValArgGlnThrAspGlnIleLeuIleGlu 113  
Qy 571 ATCTGGCGCAAGAGCGCATCGGACCGTGTGATCGCGCCCGCCCGCTGAAATCATC 630  
Db 114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133  
Qy 631 GGCGCGACATCTGACCCAGCTGGCTGGCTGACCTGAACTTCCCATCAGCCCATCGAG 690  
Db 134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153  
Qy 691 ACCGTGCGCGTGAAGCGCGCATGACCGCGCCCAAGGTGAAGCATGTGGCCCTG 750  
Db 154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnIleProLeu 173  
Qy 751 ACCGAGGAGAGATCAAGCGCTGACCGCCATCTCGCGAGGAGATGGAGAGGAGCGCAAG 810  
Db 174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluLys 193  
Qy 811 ATCACCAGATCGCGCGCGAGAACCCCTAACAACCCCGCTGTTGCGCATCAAGAAGAG 870  
Db 194 IleSerLysIleGlyProGluAsnProTrpAsnThrProValPheAlaIleLysLys 213  
Qy 871 GACAGCACCAAGTGGCGAGCTGGTGAATCTTCGCGAGCTTGAACAAGCGCCACCGAGAC 930  
Db 214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233  
Qy 931 TTCTGGAGGTGACGTGGGCGATCCCCACCGCGCGCTGAAAGAGAGAGAGCGGTG 990  
Db 234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 253  
Qy 991 ACCGTGTGAGCGTGGGCGACGCTACTTACGTGCGCTGGAGCGAGCATCTCCGCAAG 1050  
Db 254 ThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluSerPheArgLys 273  
Qy 1051 TACACCGCTTCAACATCCCGCATCAACAACGAGACCCCGCGCATCCGCTACCGATAC 1110  
Db 274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGlnTrp 293  
Qy 1111 AACGTGTGCGCCAGGCGTGAAGGCGAGCGCCAGCATCTTCACAGCAGCATCACCAAG 1170  
Db 294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313  
Qy 1171 ATCTGTGAGCGCTTCCGCGCGCGCAACCCGAGATCGTATCTAACAG-----GCCGCC 1224  
Db 314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTrpGlnTrpMetAspAsp 333  
Qy 1225 CTGTACCTGGCGACGACCTGGAGATCGCGCAGCACCGCGCCAGATCGAGGAGCTCGGC 1284  
Db 334 LeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353  
Qy 1285 AAGCACCTGTGCTGGGCTTCAACACCGCGCGCGAGAGAGACCAAGAGAGAGCGCCGCC 1344  
Db 354 AlaHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373  
Qy 1345 TTCTCTG-----CCCATCGAGCTGCACCGCGACAAAGTGGACCGTGCAGCGCATCGAGCTG 1398  
|||||

Db 374 PheLeuTrpMetGlyTrpGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393  
Qy 1399 CCCGAGAGAGAGAGCTGGACCGTGAACAGCATCCAGAAAGCTGGTGGGCAAGCTGAACCTGG 1458  
Db 394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413  
Qy 1459 GCCAGCCAGATCTACCCCGGCGATCAAGGTGGCGCAGCTGTGCAAGCTGTCTGGCGGCGCC 1518  
Db 414 AlaSerGlnIleTrpAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433  
Qy 1519 AAGCCCTGACCGACATCGTCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAAC 1578  
Db 434 LysAlaLeuThrAspIleValProLeuThrGlnGluAlaGluLeuGluGluAsn 453  
Qy 1579 CGCGAGATCTTCGCGAGCGCGTGCAGCGCTGTACTACGACCCCGACCAAGAGCTGTGTG 1638  
Db 454 ArgGluLeuLeuArgGluProValHisGlyValTrpThrAspProSerLysAspLeuVal 473  
Qy 1639 GCCGAGATCCAGAAAGCGCGCACGACGATGGACCTTACCAAGATCTTACAGAGCCCTTC 1698  
Db 474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTrpGlnGluProPhe 493  
Qy 1699 AAGAACCTGAGAGACCGCGCAAGTACGCCAAGATGCGCACCGCGCCACCAAGAGCTGAAG 1758  
Db 494 LysAsnLeuLysThrGlyLysTrpSerArgLysArgSerAlaHisThrAsnAspValArg 513  
Qy 1759 CAGCTGACCGAGCGCGTGCAGAGATCGCATCGAGAGCATCGTGTATCTGGGCAAGACC 1818  
Db 514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533  
Qy 1819 CCAGTTCGCGCTGCGCATCCAGAGGAGACCTTGGGAGACCTGGTGGACCGACTACTTGG 1878  
Db 534 ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTrp 553  
Qy 1879 CAGCGACCTGGATCCCGCGAGTGGAGTTGCGTAACACCGCGCGCTGGTGAAGCTGTGG 1938  
Db 554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573  
Qy 1939 TACCAGCTGAGAGAGAGCGCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC 1998  
Db 574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTrpValAspGlyAlaAla 593  
Qy 1999 AACCGCGAGACCAAGATCGGCAAGCGCGTACGTGACCGACCGCGCGCGCGAGAGATC 2058  
Db 594 SerArgGluThrLysLeuGlyLysAlaGlyTrpValThrAspArgGlyArgGlnLysVal 613  
Qy 2059 GTGAGCTGACCGAGACCAACCAAGACCGAGCTGCGAGCGCATCCAGCTGGCGCTG 2118  
Db 614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633  
Qy 2119 CAGGACAGCGCGCAGCGAGGTGAACATCGTACCGACAGCGAGTACGCGCTGGGCGATCATC 2178  
Db 634 GlnAspSerGlySerGluValAlaAsnIleValThrAspSerGlnTrpAlaLeuGlyIleIle 653  
Qy 2179 CAGCGCCAGCGCCAGAGAGCGAGCGAGCTGTGTAACACAGATCATCGAGCGCTGATC 2238  
Db 654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673  
Qy 2239 AAGAAGAGAGAGGTGTACTGTAGCTGGGTGCGCGCGCGCGCGAGCGCGCGAGAG 2298  
Db 674 LysLysGluLysValTrpLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693  
Qy 2299 CAGATCGACAGCTGGGTGAGCAAGGCGCATCCGAAAGGTGTCTTCTGGAGCGCGCATCGAT 2358  
Db 694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713  
RESULT 12  
US-10-296-734-1470  
; Sequence 1470, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A



594	Ser	Arg	Glu	Thr	Lys	Leu	Gly	Lys	Ala	Gly	Tyr	Val	Thr	Asp	Arg	Gly	Arg	Gln	Lys	Val	613
2059	GTG	AGC	CTG	AGC	AGC	AGC	ACC	CAAC	CCAGA	AGAC	CGAG	CTGC	GAGG	CCAT	CTC	GAG	TGG	CGC	CTG	2118	
614	Ile	Ser	Leu	Thr	Glu	Thr	Thr	Asn	Gln	Lys	Thr	Glu	Leu	His	Ala	Ile	His	Leu	Ala	Leu	633
2119	CAG	GAC	AGC	CGC	GAG	GTT	GAA	CTG	TAC	CCG	AC	CGAC	CGAC	GAT	CGC	CTG	GCG	ATC	ATC	2178	
634	Gln	Asp	Ser	Gly	Ser	Glu	Val	Asn	Ile	Val	Thr	Asp	Ser	Gln	Tyr	Ala	Leu	Gly	Ile	653	
2179	CAG	CCC	CCG	CCC	CAAG	CGAG	CGAG	CTG	TGT	GAA	CCAG	AT	CTC	GAG	CGA	CTC	ATC	ATC	2238		
654	Gln	Ala	Gln	Pro	Asp	Arg	Ser	Glu	Ser	Glu	Val	Ser	Gln	Ile	Ile	Glu	Glu	Leu	Ile	673	
2239	AAG	AAG	CAG	AGG	TGT	ACT	TGAG	CTG	TGG	TCC	CCG	CCCA	AGG	CGAT	CTG	CGG	CGA	ACG	AG	2298	
674	Lys	Asp	Glu	Lys	Val	Tyr	Leu	Ser	Trp	Val	Pro	Ala	His	Lys	Gly	Ile	Gly	Lys	Glu	693	
2299	CAG	ATC	GAC	CAAG	CTG	TGT	GAG	CAAG	GGG	CAT	CCG	CAAG	CTG	TGT	TCT	TGG	ACG	CGC	ATC	2358	
694	Gln	Val	Asp	Lys	Leu	Val	Ile	Ser	Gly	Ile	Arg	Lys	Val	Leu	Phe	Leu	Asp	Gly	Ile	713	

## RESULT 13

```

US-10-296-734-1471
; Sequence 1471, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules an
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; CURRENT FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1471
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: POL Consensus A
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (15)..(15)
; OTHER INFORMATION: X = unknown amino aci
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (22)..(23)
; OTHER INFORMATION: X = unknown amino aci
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (26)..(26)
; OTHER INFORMATION: X = unknown amino aci
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (30)..(30)
; OTHER INFORMATION: X = unknown amino aci
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (36)..(36)
; OTHER INFORMATION: X = unknown amino aci
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (38)..(39)
; OTHER INFORMATION: X = unknown amino aci
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (41)..(44)
; OTHER INFORMATION: X = unknown amino aci
; FEATURE:
; NAME/KEY: MISC_FEATURE

```

```

; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (335)..(335)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (365)..(365)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (369)..(369)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (373)..(373)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (402)..(403)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (434)..(434)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (439)..(439)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (452)..(452)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (470)..(470)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (480)..(480)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (487)..(487)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (514)..(514)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (534)..(535)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (544)..(544)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (553)..(553)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (593)..(593)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (610)..(610)
; OTHER INFORMATION: X = unknown amino acid

```

```

; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (675)..(675)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:

Alignment Scores:
Pred. No.:      3.55e-148      Length:      1006
Score:          3208.50        Matches:      614
Percent Similarity: 88.26%      Conservative: 25
Best Local Similarity: 84.81%    Mismatches:  74
Query Match:     70.44%        Indels:      11
DB:              15           Gaps:        4

US-09-610-313B-32 (1-2457) x US-10-296-734-1471 (1-1006)

QY 220 TTCTTCGGCAGAGACCTGGCCCTTCCCGCAGGCGAGCCCGCGAGTTCCCGCAGCGAGCAG 279
Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArg***PheSerSerGluGln 20
QY 280 AACCGCGCCAAACAGCCCGCAGCGCGAGCTGCAGGTGCGCGGCGACCAACCC 333
Db 21 Thr*****AsnSer***ThrSerArg***LeuTrpAspGlyGly***Asp*****Leu 40
QY 334 GCGACGAGGCGCGCGCGCGCGCGCGCGC-----ACCTGAACCTTCGCC 378
Db 41 *****Gly***Glu***GlnGly***Gly*****SerPhe***PhePro 60
QY 379 CAGATCACCTGTGGCAGCGCCCTGTGTAGCATCAAGTGTGGCGCGCAGATCAAGGAG 438
Db 61 GlnIleThrLeuTrpGlnArgProLeuValThrVal***Ile***GlyGlnLeuIleGlu 80
QY 439 GCCCTGTGGACACCGCGCGCGAGCACACCTGCTGGAGGAGATGAGCTGCCCGCAAG 498
Db 81 AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLys 100
QY 499 TGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCAGTACAC 558
Db 101 TrpLysProLys***IleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGln 120
QY 559 ATCTGTATCGAGATCTCGGCAAGAGCCATCGGCACCTGCTGCTGATCGGCCCGCACCC 618
Db 121 IleLeuIleGluIleCysGlyLys***IleGlyThrValLeuValGlyProThrPro 140
QY 619 GTGACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTCACTTCCCTCCATC 678
Db 141 ValAsnIleIleGlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160
QY 679 AGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGAGCGCGCCCAAGGTGAAG 738
Db 161 SerProIleGluThrValProValLysLeuLysPro***MetAspGlyProLysValLys 180
QY 739 CAGTGGCCCTTGACCGAGGAGAAGATCAAGCCCTGACCGCCATCTCGAGGAGATGAG 798
Db 181 GlnTrpProLeuThrGluGluLysIleLysAlaLeuThr***IleCys***GluMetGlu 200
QY 799 AAGGAGGCAAGATCACCAGATCGGCGCGCGAGAACCCCTACACACCCCGGTGTCGCC 858
Db 201 LysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAla 220
QY 859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGAGCTCCGCGAGCTGAACCAAG 918
Db 221 IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 240
QY 919 CGCACCCAGGACTTCTGGGAGGTGCAGTGGGCAATCCCGCCACCCCGCGGTGGAAGAG 978
Db 241 ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis***AlaGlyLeuLysLys 260
QY 979 AAGAAGCGGTGACCGTCTGAGCGTGGCGACCGCTACTTCAGCGTCCCGCTGACGAG 1038
Db 261 LysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAsp*** 280
QY 1039 GACTTCCGCAAGTACACCGCCTTCACCATCCCGCATCAACACGAGACCCCGCGCATC 1098

```

Db 281 \*\*\*\*PheArgLysTyrThrAlaPheThrIleProSer\*\*\*\*AsnAsnGluThrProGly\*\*\* 300  
Qy 1099 CGCTACCAAGTACACGCTGCTGCCAGGCTGGAAGGGCCAGCCCGACCATCTTCCAGAGC 1158  
Db 301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro\*\*\*IlePheGln\*\*\* 320  
Qy 1159 AGCATGACCAAGATCTCTGGAGCCCTTCCGGCCCGCAACCCCGAGATCGTGATCTACAG 1218  
Db 321 SerMetThrLysIleLeuGluProPheArg\*\*\*\*\*Pro\*\*\*IleValIleTyrGln 340  
Qy 1219 -----GCCCCCTGTACGTGGGCGAGCTGGAGATCGCCAGCACCGGCCAGATC 1272  
Db 341 TyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360  
Qy 1273 GAGGAGCTGGCGAAGCACCTGCTCGCTGGGCTTCCACCCCGCCAGCAAGAGACCCAG 1332  
Db 361 GluGluLeuArg\*\*\*HisLeuLeu\*\*\*TrpGlyPhe\*\*\*ThrProAspLysLysHisGln 380  
Qy 1333 AAGAGCCCTCTCTCTG-----CCCATCGAGCTGCACCCCGACCAAGTGGACCGTGCAG 1386  
Db 381 LysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTrpThrValGln 400  
Qy 1387 CCCATCGAGCTGCCGAGAGGAGCTGCACCGTGAACGATCCAGAGCTGTGGC 1446  
Db 401 Pro\*\*\*\*\*LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420  
Qy 1447 AAGCTGAACCTGGGCGACCATCTACCCCGCATCAAGTGCAGCCAGCTGTGCAAGCTG 1506  
Db 421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys\*\*\*LysGlnLeuCys\*\*\*Leu 440  
Qy 1507 CTGGCGGCGCCAAAGCCCTGACGACATCTGCTGCCCTGACGAGGAGCGGAGCTGAG 1566  
Db 441 LeuArgGlyAlaLysAlaLeuThrAspIleVal\*\*\*LeuThrGluGluAlaGluLeuGlu 460  
Qy 1567 CTGGCGGAGACCGGAGATCTCGCGAGCCGCTGCACCGCTGTACTAGACCCCGCAGC 1626  
Db 461 LeuAlaGluAsnArgGluIleLeuLys\*\*\*ProValHisGlyValTyrAspPro\*\*\* 480  
Qy 1627 AAGCACTGTGTGGCGGAGATCTCAGAACGAGCGGCCAGCACAGTGGACCTACCATGCTAC 1686  
Db 481 LysAspLeuValAlaGlu\*\*\*GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500  
Qy 1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCACCGCCACACC 1746  
Db 501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla\*\*\*LysArgSerAlaHisThr 520  
Qy 1747 AAGCACTGAAGCACTGACCGAGCGCTGCAGAGATCGCCATGAGAGCATCTGTATC 1806  
Db 521 AsnAspValLysGlnLeuThrGluValAlaGlnLysVal\*\*\*\*\*GluSerIleValIle 540  
Qy 1807 TGGGGCAAGACCCCAAGTTCCGCTCCCATCCAGAAAGAGACCTGGGAGACTGTGTGG 1866  
Db 541 TrpGlyLys\*\*\*ProLysPheArgLeuProIleGln\*\*\*GluThrTrpGlu\*\*\*TrpTrp 560  
Qy 1867 ACCGACTACTGGCAGGCCACCTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTG 1926  
Db 561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580  
Qy 1927 GTGAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCCGAGACTTCTAGCTG 1986  
Db 581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle\*\*\*GlyAlaGluThrPheTyrVal 600  
Qy 1987 GACGGCGCCCAACCGCAGAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGGGC 2046  
Db 601 AspGlyAlaAlaAsnArgGluThrLys\*\*\*GlyLysAlaGlyTyrValThrAspArgGly 620  
Qy 2047 CGGCAGAAAGTCTGTGACCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGGCCATC 2106  
Db 621 ArgGlnLysValValSerLeuThrGluThrAsnGlnLysThrGluLeuHisAlaIle 640  
Qy 2107 CAGCTGGCCCTGACGAGCACCGGAGGAGGTGAACATCTGTGACCGACACCGATGAGCC 2166

Db 641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660  
Qy 2167 CTGGGATCATCCAGGCCCGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2226  
Db 661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu\*\*\*ValAsnGlnIleIle 680  
Qy 2227 GAGCAGCTGATCAAG 2286  
Db 681 GluLysLeuIle\*\*\*Lys\*\*\*LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700  
Qy 2287 GGCGGCAACAGCAGCAGATCGAAGCTGTGAGCAAGAGAGAGAGAGAGAGAGAGAG 2346  
Db 701 GlyLysAsnGlnGlnValAspLysLeuValSer\*\*\*GlyIleArgLysValLeuPheLeu 720  
Qy 2347 GACGCGATCGAT 2358  
Db 721 AspGlyIleAsp 724

RESULT 14  
US-10-301-661A-6  
; Sequence 6, Application US/10301661A  
; Publication No. US20030157660A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-  
; APPLICANT: INSERM  
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: MAUCLERE, Philippe  
; APPLICANT: LOUSSERT-AJAKA, Ibtissem  
; APPLICANT: SIMON, Francois  
; APPLICANT: SARAGOSTI, Sentob  
; APPLICANT: BARRE-SINOSSI, Francoise  
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.  
; FILE REFERENCE: 598US12  
; CURRENT APPLICATION NUMBER: US/10/301,661A  
; PRIOR FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US/09/319,588C  
; PRIOR FILING DATE: 1999-08-27  
; PRIOR APPLICATION NUMBER: FR96/15087  
; PRIOR FILING DATE: 1996-12-09  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1014  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-10-301-661A-6

Alignment Scores:  
Pred. No.: 7,22e-145 Length: 1014  
Score: 3140.50 Matches: 579  
Percent Similarity: 89.70% Conservative: 74  
Best Local Similarity: 79.53% Mismatches: 60  
Query Match: 68.95% Indels: 15  
Gaps: 5

US-09-610-313B-32 (1-2457) x US-10-301-661A-6 (1-1014)

Qy 220 TTCTTCGGGAGAGACTGCGCTTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 279  
Db 1 PhePheArgGluGluGluLeuValSerLeuGluArgGluThrArgLysLeuProAspAsn 20  
Qy 280 AAC-----CGCCCAACAGCCCGCCAGCGCGAGCTGCAGGTG----- 318  
Db 21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40  
Qy 319 ---CGCGGCGACAACCCCGCAGAGCGCGCGCCAGCGCCAG-----GGCACC 366  
Db 41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60  
Qy 367 CTGAACCTTCCCGAGATCATCCCTGTGCGAGCGCCCTGTGAGCATCAAGTGGGGCGC 426  
Db 61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80



QY 427 CAGATCAAGGAGCCCTGCTGACACCGCGCCGACACCGCTGCTGGAGGATGAGC 486  
Db 81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluLeuGln 100  
QY 487 CTGCCCGGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCAATCAAGTGGCC 546  
Db 101 LeuGluGlyLeuTrpLysProLysMetIleGlyGlyIleGlyPheIleLysValArg 120  
QY 547 CAGTACGACCATCTGATCAGATCTGCGGCAAGAGGCGCATCGGCAACGCTGCTGATC 606  
Db 121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140  
QY 607 GCGCCCAACCCCGTGAACATCATCGCGCGCAACATCTGACCCACTGGCTGGCCCTG 666  
Db 141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160  
QY 667 AACTTCCCATCAGCCCATCAGACCGTCCCGTGAAGCTGAAGCCCGGATGACGCGC 726  
Db 161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180  
QY 727 CCCAAGTGAAGCATGCGCCCTGACGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
Db 181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200  
QY 787 GAGGAGATGGAGAGGCGCAAGATCACCAGATCGCGCCCGCCGAGAACCCCTACACACC 846  
Db 201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220  
QY 847 CCGCTGTTCCGCATCAAGAAGAGACACACCAAGTGGCGCAAGCTGGTGGACTTCGCG 906  
Db 221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240  
QY 907 GAGCTGAACACCGCACCCAGACTTCTGGAGAGTGCAGCTGGGATCCCCACCCCGCC 966  
Db 241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260  
QY 967 GCGCTGAAGAGAGAGAGCGTACCGTCTGGAGCGTGGCGCGCTACTTCTACGCTG 1026  
Db 261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280  
QY 1027 CCGCTGGACGAGGACTTCGCAAGTACACCGCTTCCACATCCCGCATCAACACGAG 1086  
Db 281 ProLeuAspLysAspPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGlu 300  
QY 1087 ACCCGCGGATCCGCTACAGTACAGTACAGTCTGCGCCCGGCTGGAGGCGAGCCCGCAG 1146  
Db 301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320  
QY 1147 ATCTTCCAGACGATGACCAAGATCTGGAGCCCTTCCGCGCCGCAACCCCGAGATC 1206  
Db 321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysPheProGluIle 340  
QY 1207 GTGATCTACCG- - - - -GCCCGCTGTACGTGGGCGAGCGCTGGAGATCGCCGAGCAG 1260  
Db 341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360  
QY 1261 CGCGCCCAAGATCGAGCTGCGCAAGCATCTGCTGCGCTGGGCTTCAACACCCCGCAG 1320  
Db 361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380  
QY 1321 AAGAGGACCAAGGAGCCCGCTTCTCTG- - - - -CCCATCGAGCTGACCCCGCAGAG 1374  
Db 381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400  
QY 1375 TGGACCGTGCAGCCCATCGAGTGGCCGAGAGAGAGTGGACCGCTGAACACATCCAG 1434  
Db 401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420  
QY 1435 AAGCTGGTGGCAAGCTGAACCTGGCGCGCCAGATCTACCCCGGATCAAGGTGGCGCAG 1494  
Db 421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440

QY 1495 CTGTGCAAGTGTGCTGGCGCGCCCAAGCCCTGACCGACATCTGTGCTGACCCGAGGAG 1554  
Db 441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460  
QY 1555 CCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTTCCGCGAGCCCGTGCACGCGCTGTAC 1614  
Db 461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480  
QY 1615 TACGACCCCAACAGGACCTGTGGCCGAGATCCAGAGACGAGGCGCACGACGAGTGACC 1674  
Db 481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnGlyValTrpThr 500  
QY 1675 TACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCAAGATGCGC 1734  
Db 501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520  
QY 1735 ACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAG 1794  
Db 521 SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu 540  
QY 1795 AGCATCTGATCTGGGCGCAAGACCCCAAGTTCGCTGCCCTGCCATCCAGAGGAGACCTGG 1854  
Db 541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560  
QY 1855 GAGACCTGTGTCAGCCACTACTTGGCAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1914  
Db 561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580  
QY 1915 ACCCGCCCTGTGTAAGCTGTGTACAGCTGGAGAGAGGCCATCATCGCCCGCAG 1974  
Db 581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600  
QY 1975 ACCTTCTACGTGGAGCGCGCCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTG 2034  
Db 601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620  
QY 2035 ACCGACCGCGCGCGCAGAGATCTGAGCTGACCGAGACCCACCAACAGAGAGCCGAG 2094  
Db 621 ThrAspArgGlyArgGlnLysValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640  
QY 2095 CTGACGAGCATCCAGCTGGCCCTGCGAGCACGCGCAGCGAGGTGAACATCGTGACCGAC 2154  
Db 641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660  
QY 2155 AGCCAGTACGCGCTGGGATCATTCAGCGCCCGCAGCGCCGAGCGAGCGAGCGTGTG 2214  
Db 661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680  
QY 2215 AACAGATCATCGAGCAGCTCATCAAGAGAGAGAGGTGTACCTGAGTGGTGGCCCGCC 2274  
Db 681 SerGlnIleIleGluGluLeuLysLysGluArgValTyrLeuSerTrpValProAla 700  
QY 2275 CACAAGGCGATCGCGCGCAACGAGCAGATCGCAAGCTGGTGGAGCAAGGCGATCCGCAAG 2334  
Db 701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720  
QY 2335 GTGCTGTTCTCGAGCGCATCATGAT 2358  
Db 721 IleLeuPheLeuAspGlyIleGlu 728  
RESULT 15  
US-09-952-060-35  
; Sequence 35, Application US/09952060  
; Publication No. US2003004421A1  
; GENERAL INFORMATION:  
; APPLICANT: Emini, Emilio A.  
; APPLICANT: Youil, Rima  
; APPLICANT: Bett, Andrew J.  
; APPLICANT: Chen, Ling  
; APPLICANT: Kaslow, David C.  
; APPLICANT: Shiver, John W.  
; APPLICANT: Toner, Timothy J.  
; APPLICANT: Casimiro, Danilo R.

Qy	13	ATGCGCGAGGCCATGAGCCAGCCACC---	AGCGCCAAACATCTGTATGACGCGCAAC	69
Db	363	LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn	382	
Qy	70	TTCAAGGGCCCCAAGCGGATCATCAAGTGCTTCAACTGCGCAGGAGGAGGCGCCATCGGC	129	
Db	383	PheArgAsnGlnArgGlyThrValIysCysPheAsnCysGlyIysValGlyIHisIleAla	402	
Qy	130	CGCAACTGCGCGCCCCCGCAAGAAGGGCTGCTGGAAGTGCGGCAAGAGGAGGCGCCACAG	189	
Db	403	LysAsnCysArgAlaProArgIysIysGlyCysItrIysCysGlyIysGluGlyHisGln	422	
Qy	190	ATGAGGACTGCACGAGGCGCAGGCCAACTTCTTCGCGAGAGACTGGCTTCCCCCAG	249	
Db	423	MetLysAspCysAsnGluArgGlnAlaAsnPheLeu-GlyIysIleItrpProSerHisIy	442	
Qy	250	GGCAGGGCCCGGAGTTCCTCCAGCGAGCAGAACCGCGCCCAACAGCCCCACCGCCGAG	309	
Db	442	sgIysArgProGlyAsnPheLeuGlnIserArgProGluProThrAlaProProGluGluse	462	
Qy	310	CTGCAAGTGCGGC-----GGCACAAACCCCGCAGCGAGCGCGCGCGCAGCCCGAGGGC	363	
Db	462	rPheArgPheGlyGluGluIysThrThrProSerGlnIysGlnIuproIleAspIys---	481	
Qy	364	ACCTGAATCTCCCCAGATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGTGGGCG	423	
Db	482	GluLeu-----TyrProLeuAlaSerLeuArg-----	490	
Qy	424	GGCCGATCAAGGAGCCCTGCTGCACACCGGCGCCGACCGCTGTGGAGGAGATG	483	
Db	490	-----	490	
Qy	484	AGCCTGCCCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGCGCTTCATCAAGGTG	543	
Db	490	-----	490	
Qy	544	CGCCAGTACGACAGATCCTGTAGTGAATCTGCGGCAAGAGGCCCATCGGCACCGTGCTG	603	
Db	490	-----	490	

QY 1672 ACCTACAGATCTACGAGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATG 1731  
 Db |||||  
 840 ThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 1732 CGCACCGCCACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCCCATG 1791  
 Db |||||  
 860 ArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr 879  
 QY 1792 GAGACATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAGGAGACC 1851  
 Db |||||  
 880 GluSerIleValIleTyrGlyLysThrProLysPheLysLeuProIleGlnLysGluThr 899  
 QY 1852 TGGGAGACCTGTGTGACCGACTACTGCGAGCGACCTGATCCCCGAGTGGGAGTTCGTG 1911  
 Db |||||  
 900 TrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVal 919  
 QY 1912 AACACCCCCCTCGTGAGCTGTGTACGAGCTGGAGAGGAGCCCATCATCGCGGCC 1971  
 Db |||||  
 920 AsnThrProLeuValLysLeuTyrTrpGlnLeuGluLysGluProIleValGlyAla 939  
 QY 1972 GAGACCTTCTAGCTGAGCGCGCCCAACCGCAGAGACCAAGATCGGCAAGSCGCGCTAC 2031  
 Db |||||  
 940 GluThrPheTyrValAlaGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyr 959  
 QY 2032 GTGACCGACCGCGCGCGCAGAGATCGTGAGCCTGACCGAGACCAACCAAGAGACC 2091  
 Db |||||  
 960 ValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThr 979  
 QY 2092 GAGTGCAGGCGCATCCAGCTGCGCCCTGCAGGACAGCGGCGAGGAGGTGAACATCGTGACC 2151  
 Db |||||  
 980 AlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThr 999  
 QY 2152 GACAGCCAGTACGCGCTCGGCATCATCCAGCGCCAGCCGACAGCGAGCGAGCGAGCTG 2211  
 Db |||||  
 1000 AlaSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeu 1019  
 QY 2212 GTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACTGAGCTGGGTGCC 2271  
 Db |||||  
 1020 ValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValPro 1039  
 QY 2272 GCCCAAGGGGATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGCGCATCCGC 2331  
 Db |||||  
 1040 AlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArg 1059  
 QY 2332 AAGGTGCTGTTCTCGACGCGCATCGAT 2358  
 Db |||||  
 1060 LysValLeuPheLeuAspGlyIleAsp 1069

Search completed: June 2, 2005, 05:29:26  
 Job time : 221.794 secs